# Darwins Optimierung aus der Sicht von Chemie und Physik

Peter Schuster

Institut für Theoretische Chemie, Universität Wien, Austria
and
The Santa Fe Institute, Santa Fe, New Mexico, USA


Faktum Evolution

Naturhistorisches Museum

Wien, 19.– 20.05.2009

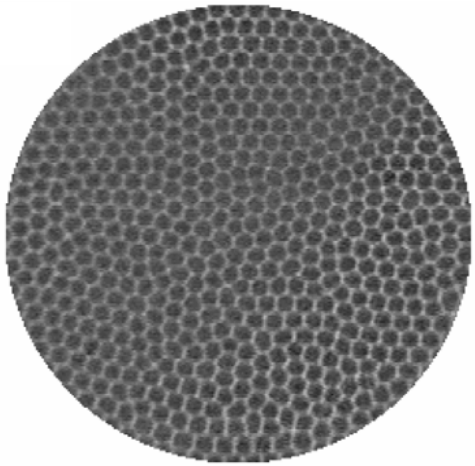
Web-Page for further information:

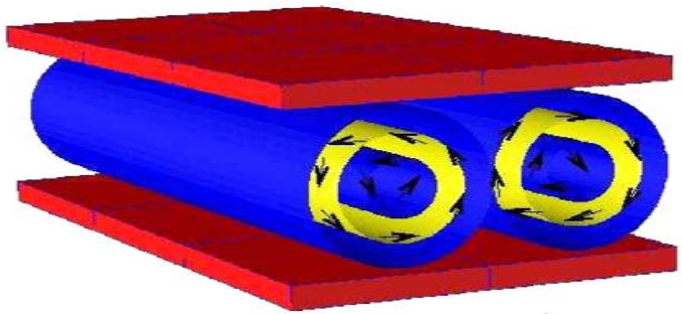
http://www.tbi.univie.ac.at/~pks

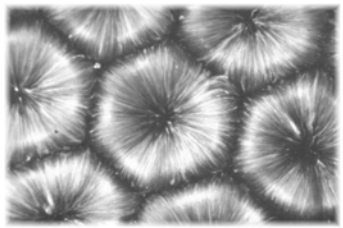
Farbmuster auf Tierfellen, -flügeln und -panzern

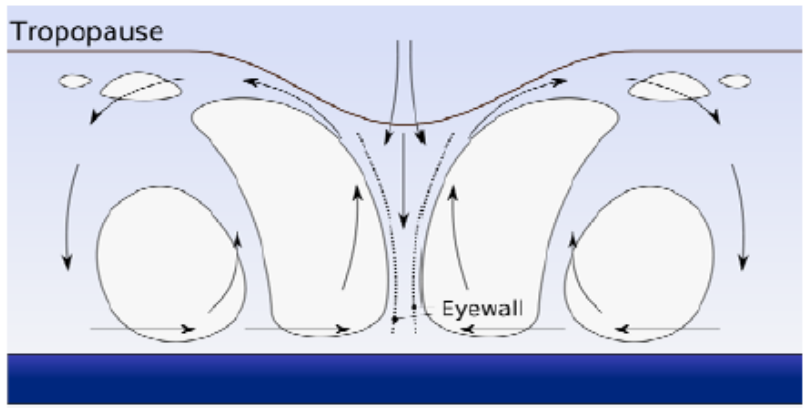
1. Musterbildung in Physik und Chemie
2. Muster in der Biologie
3. Darwins Prinzip der natürlichen Auslese
4. Vermehrung und Evolution von Molekülen
5. Chemische Kinetik der molekularen Evolution
6. Evolutionsexperimente mit Molekülen
7. Ursachen und Konsequenzen der Neutralität
8. Komplexität in der Biologie

1. **Musterbildung in Physik und Chemie**
2. Muster in der Biologie
3. Darwins Prinzip der natürlichen Auslese
4. Vermehrung und Evolution von Molekülen
5. Chemische Kinetik der molekularen Evolution
6. Evolutionsexperimente mit Molekülen
7. Ursachen und Konsequenzen der Neutralität
8. Komplexität in der Biologie

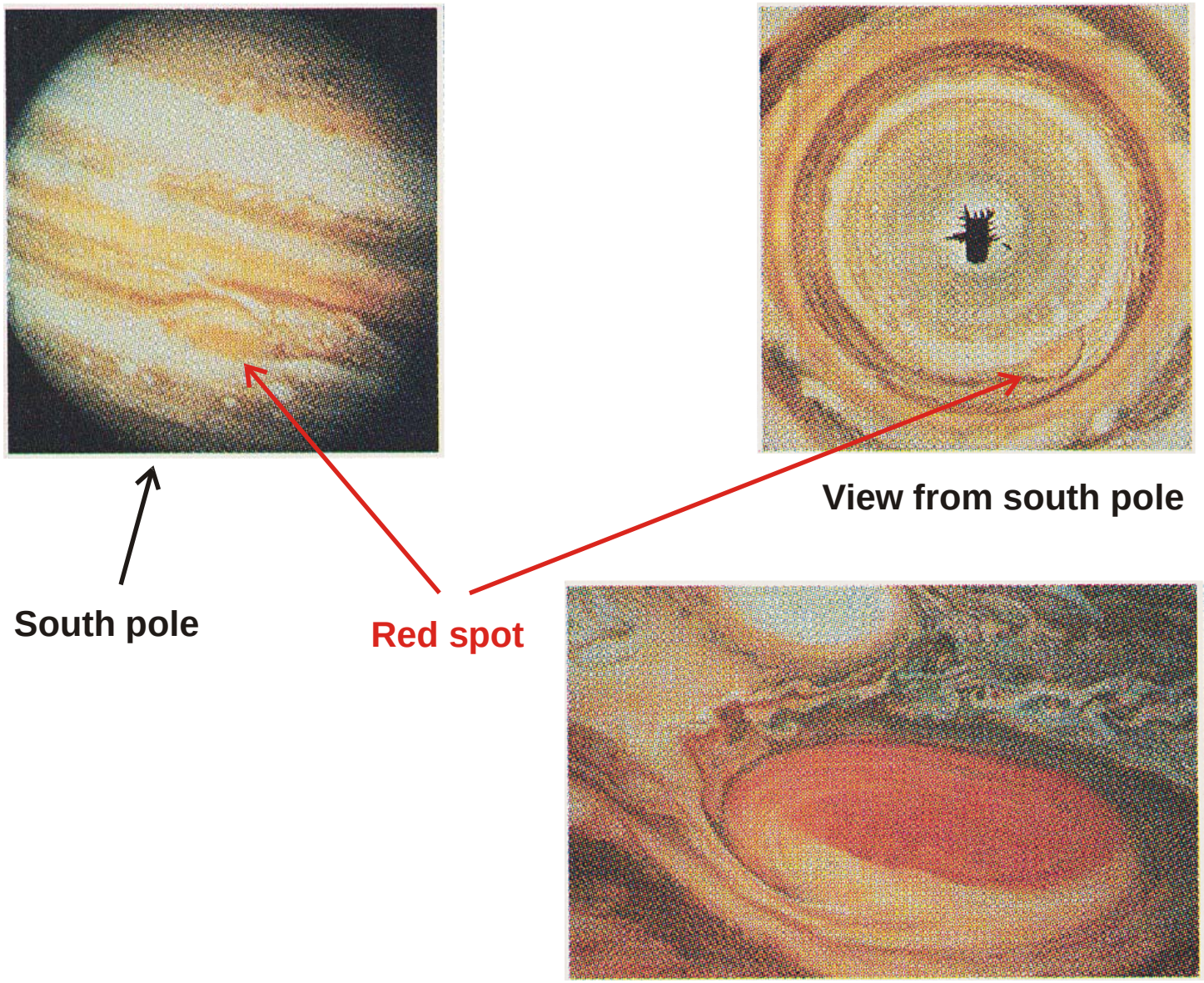







Raleigh-Bénard Konvektion und Ausbildung von Hurrikanen




Roter Fleck des Jupiters:  Beobachtung eines gigantischen Wirbels

Bilder sind entnommen von James Gleick, *Chaos*. Penguin Books, New York, 1988

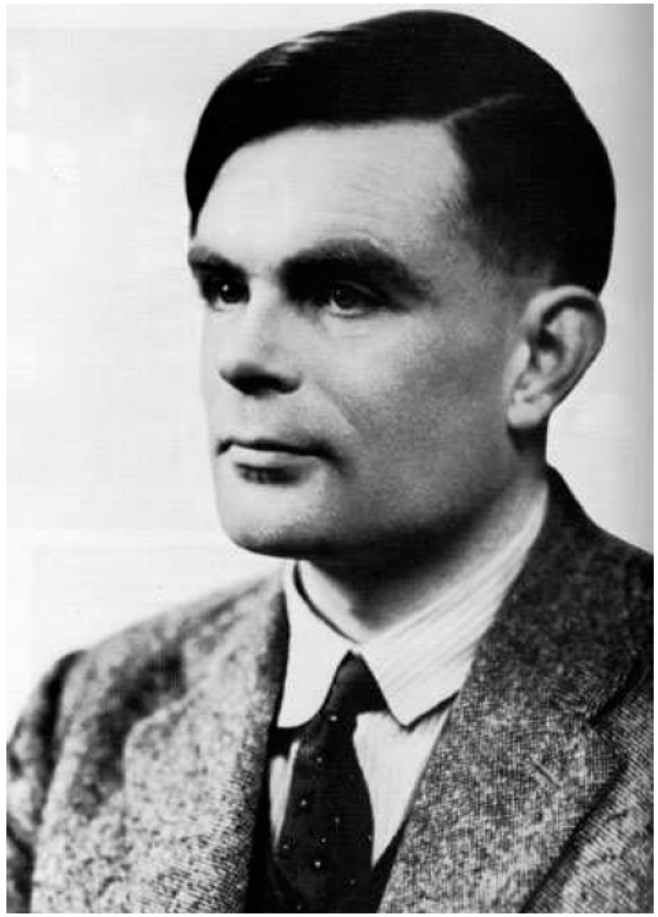
$$\frac{\partial u}{\partial t} = D_u \nabla^2 u + f(u,v)$$

$$\frac{\partial v}{\partial t} = D_v \nabla^2 v + g(u,v)$$

$$u = u(x,y,z,t) \quad \text{and} \quad v = v(x,y,z,t)$$

Veränderung in der lokalen Konzentration =

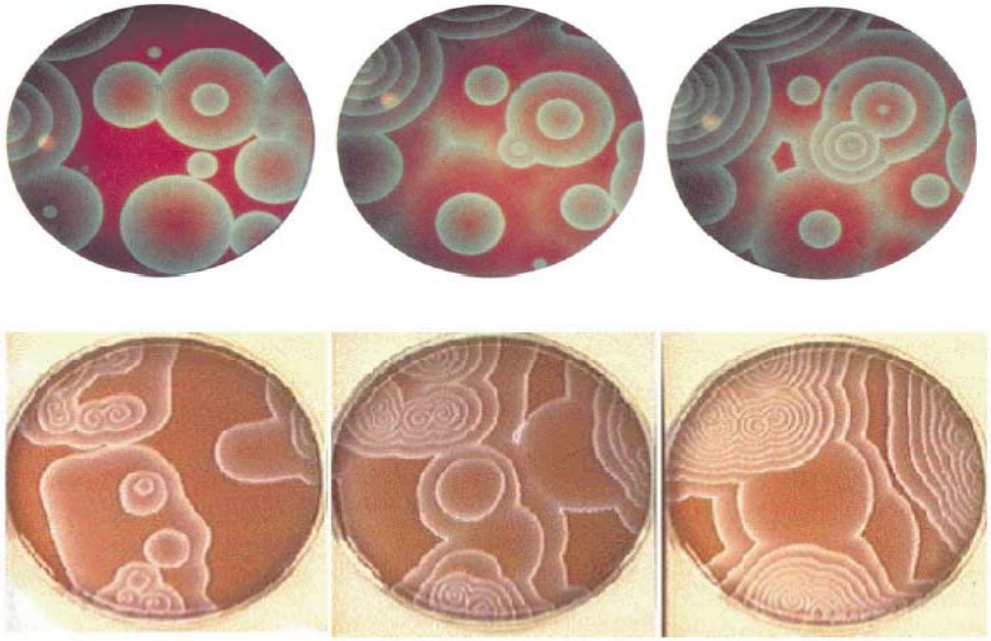
= Diffusion + Chemische Reaktion



Alan M. Turing, 1912-1954

A.M. Turing. 1952. The chemical basis of morphogenesis.
*Phil.Trans.Roy.Soc.*London B **237**:37-72.

Liesegang Ringe 1895



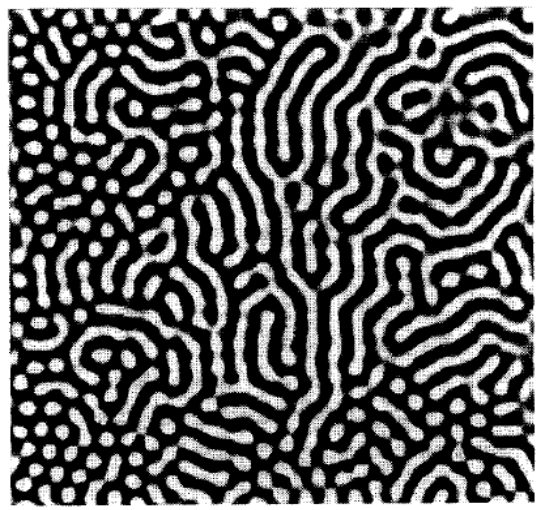
Belousov-Zhabotinskii Reaktion 1959

Musterbildung durch chemische Selbstorganisation:

Liesegang Ringe durch Fällung aus übersättigten Lösungen,

Raum-Zeit-Muster in der Belousov-Zhabotinskii Reaktion,

und stationäre Turing Muster.



Turing Muster:
Boissonade, De Kepper 1990

1. Musterbildung in Physik und Chemie
2. **Muster in der Biologie**
3. Darwins Prinzip der natürlichen Auslese
4. Vermehrung und Evolution von Molekülen
5. Chemische Kinetik der molekularen Evolution
6. Evolutionsexperimente mit Molekülen
7. Ursachen und Konsequenzen der Neutralität
8. Komplexität in der Biologie

Fellmuster in einer Inzuchtlinie von verwilderten Katzen

Eltern und Tochter

Mutter

mutmaßlicher Vater

Tochter

Bates' mimicry

Müller's mimicry

Different forms of mimicry observed in nature

milk snake

Bates' mimicry

false coral snake

coral snake

Emsley's or Mertens' mimicry

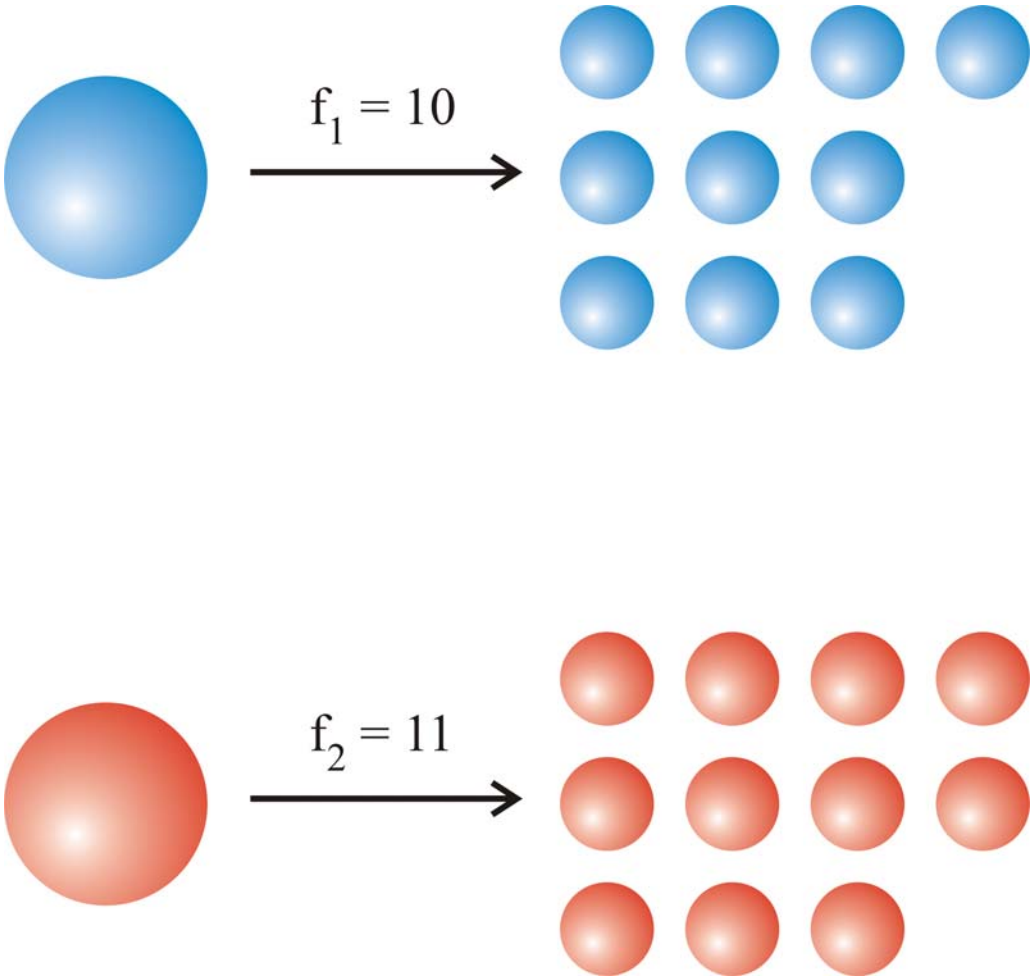
Different forms of mimicry observed in nature

1. Musterbildung in Physik und Chemie
2. Muster in der Biologie
3. **Darwins Prinzip der natürlichen Auslese**
4. Vermehrung und Evolution von Molekülen
5. Chemische Kinetik der molekularen Evolution
6. Evolutionsexperimente mit Molekülen
7. Ursachen und Konsequenzen der Neutralität
8. Komplexität in der Biologie

Drei notwendige Bedingungen für Darwinsche Evolution:

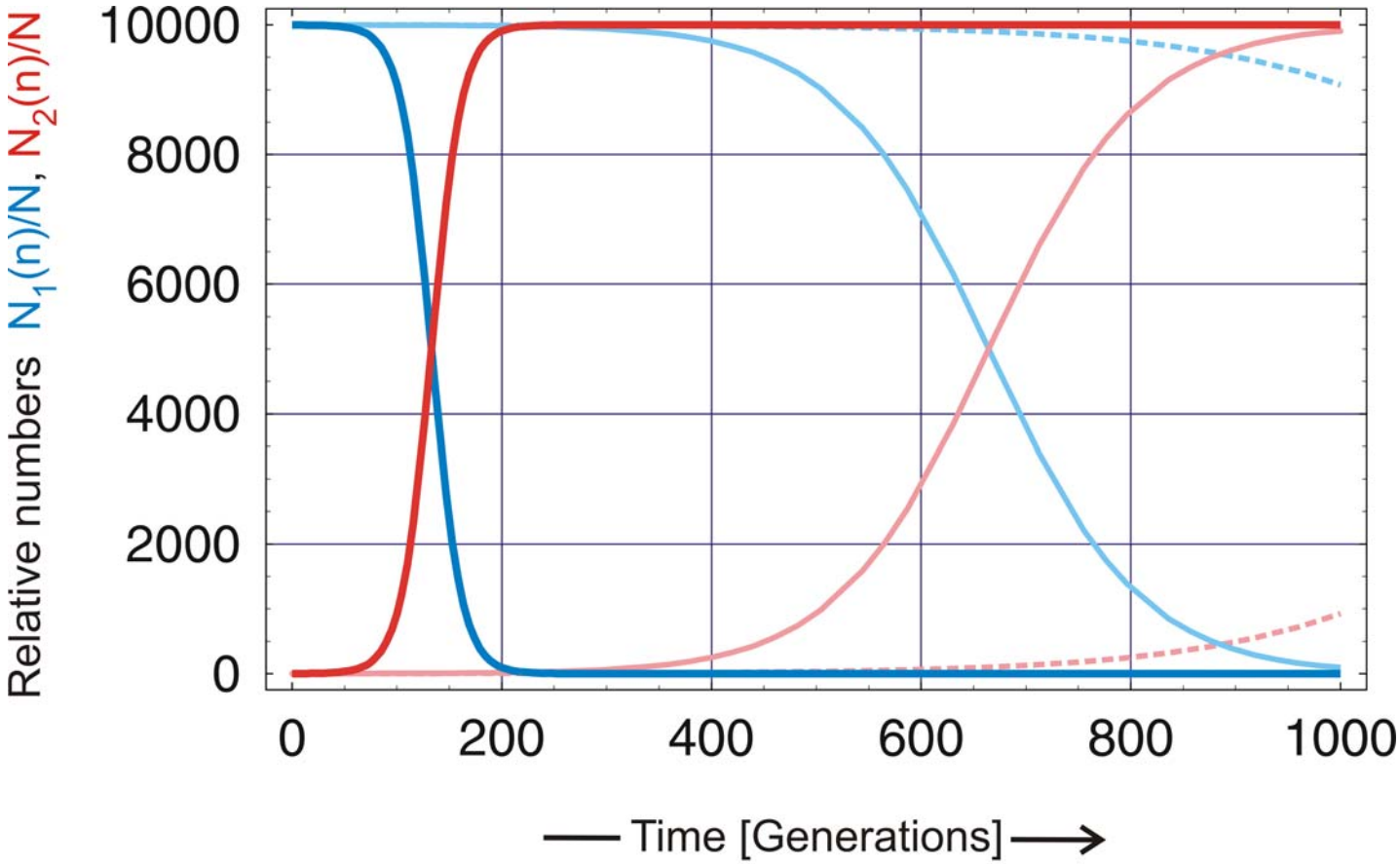
1. **Vermehrung**
2. **Variation**
3. **Selektion**

Empirisch erkanntes Prinzip der natürlichen Auslese




$$s = \frac{f_2 - f_1}{f_1} = 0.1$$

Two variants with a mean progeny of ten or eleven descendants




$$N_1(0) = 9999,\ N_2(0) = 1; \quad s = 0.1, 0.02, 0.01$$

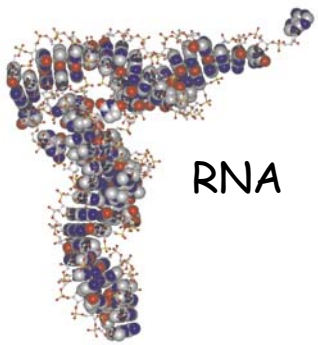
Selection of advantageous mutants in populations of N = 10 000 individuals

# Genotype, Genome

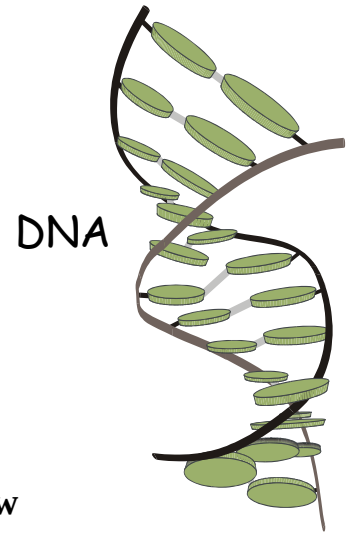
GCGGATTTAGCTCAGTTGGGAGAGCGCCAGACTGAAGATCTGGAGGTCCTGTGTTCGATCCACAGAATTCGCACCA

systems biology

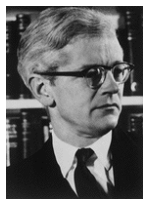
*'the new biology is the chemistry of living matter'*



RNA



DNA



John Kendrew

Phenotype

**Thomas Cech**

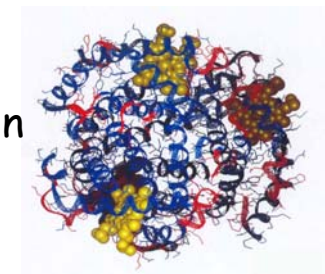
**RNA catalysis**

**Manfred Eigen**

**Linus Pauling and Emile Zuckerkandl**

**molecular evolution**

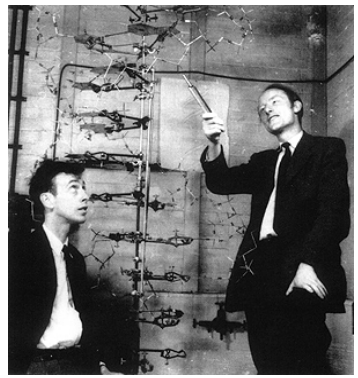
protein



**Gerhard Braunitzer**

**hemoglobin sequence**

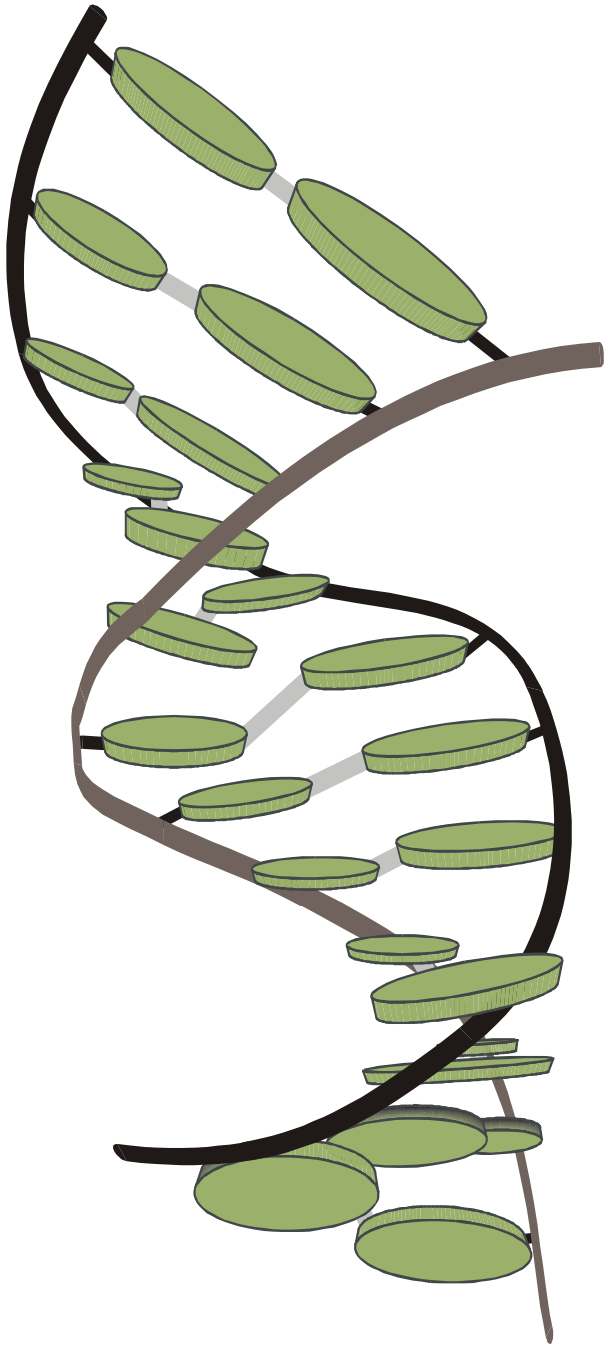
**Max Perutz**

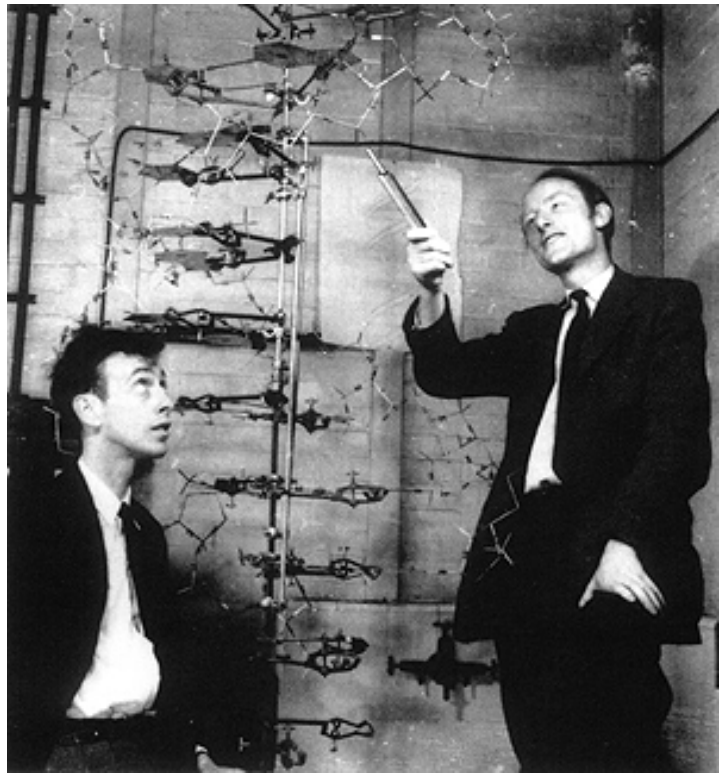


**James D. Watson und Francis H.C. Crick**

**DNA structure**

1. Musterbildung in Physik und Chemie
2. Muster in der Biologie
3. Darwins Prinzip der natürlichen Auslese
4. **Vermehrung und Evolution von Molekülen**
5. Chemische Kinetik der molekularen Evolution
6. Evolutionsexperimente mit Molekülen
7. Ursachen und Konsequenzen der Neutralität
8. Komplexität in der Biologie



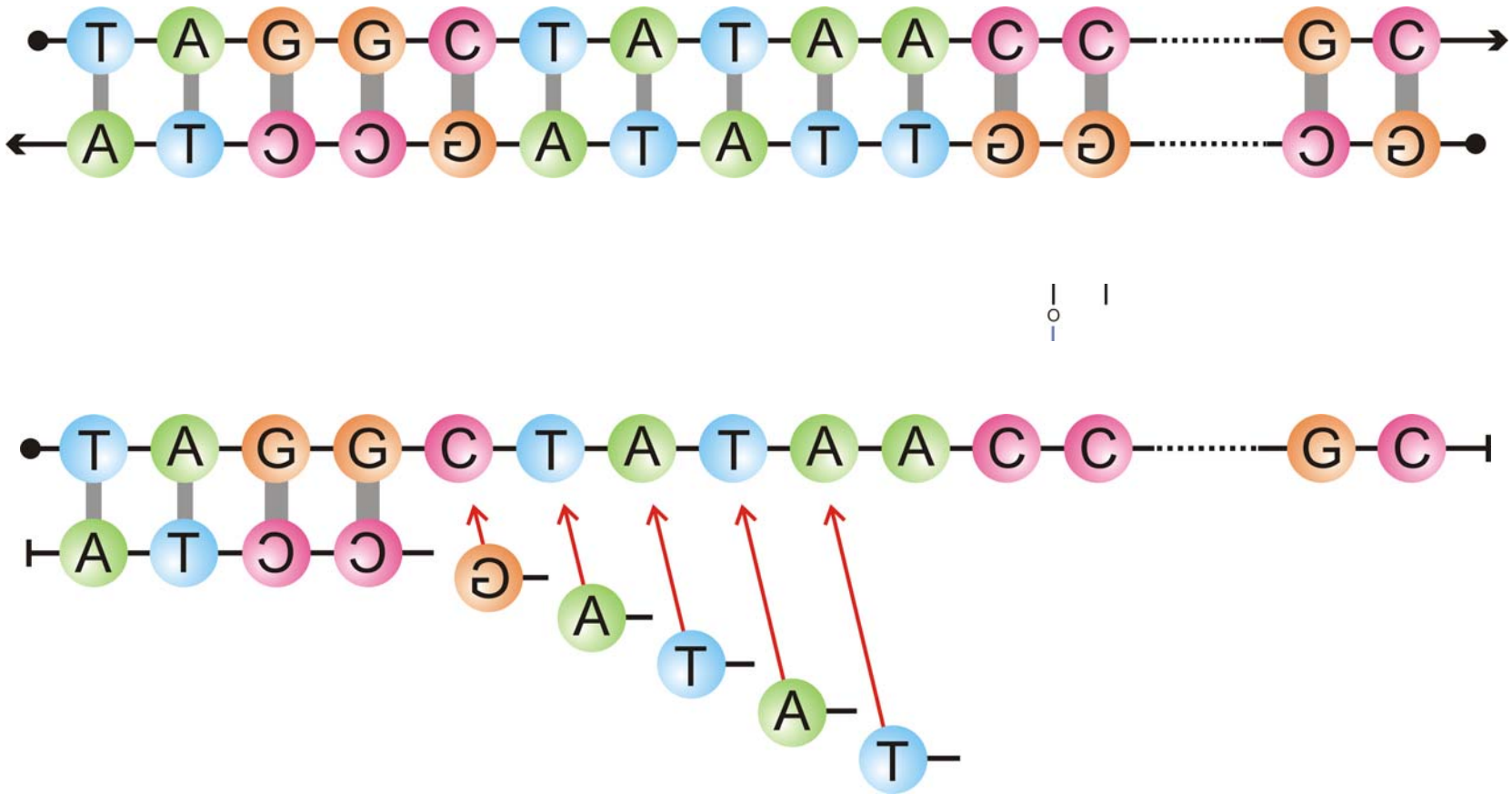


James D. Watson, 1928-, and Francis H.C. Crick, 1916-2004

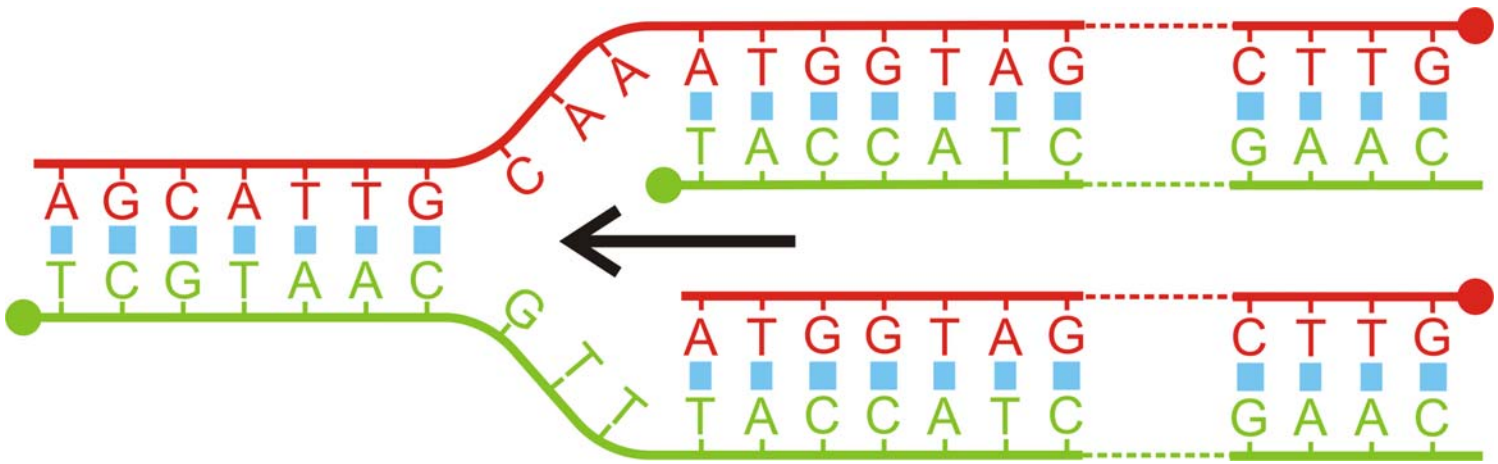
Nobel prize 1962

**1953 – 2003 fifty years double helix**

The three-dimensional structure of a short double helical stack of B-DNA

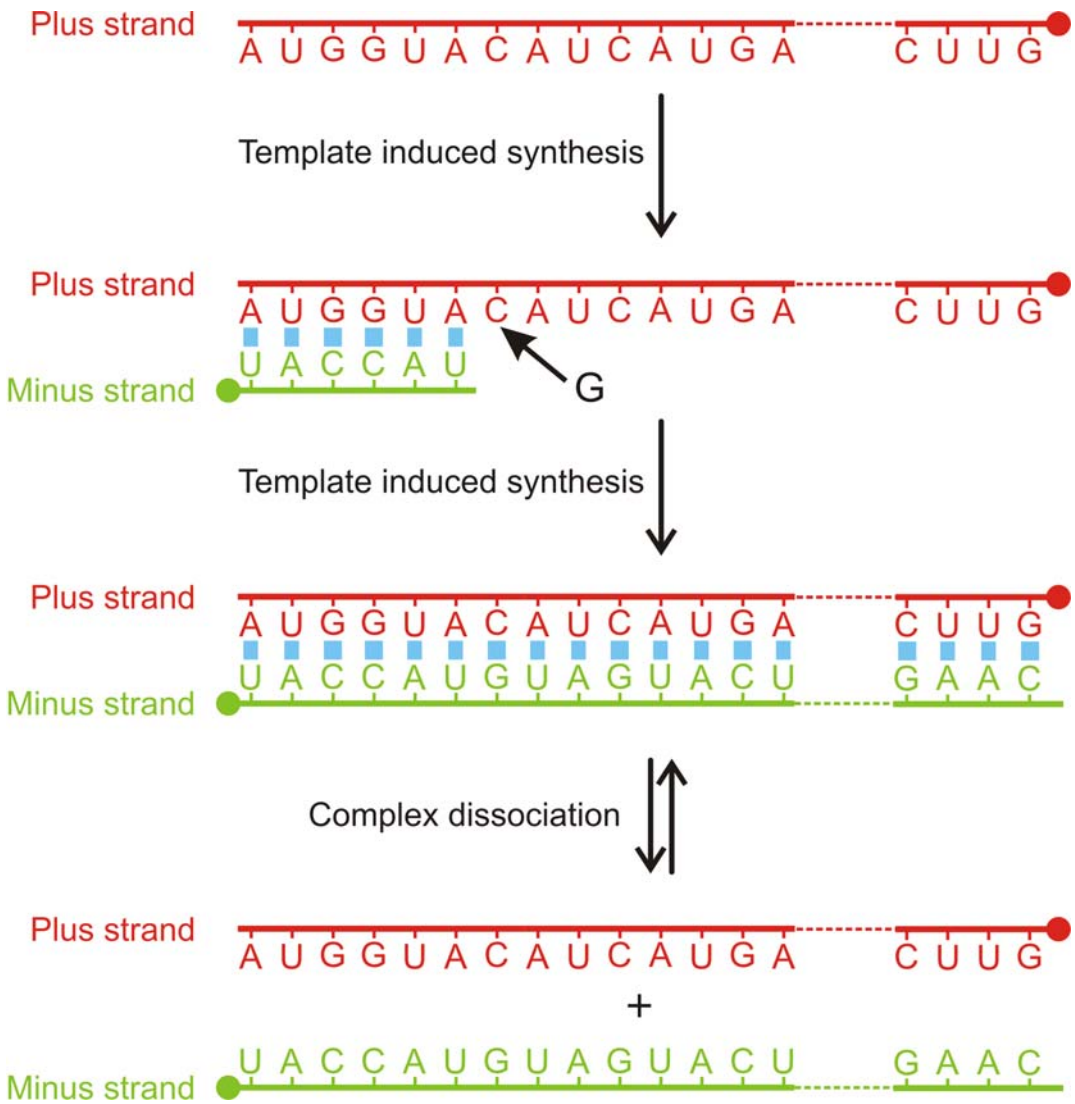



DNA structure and DNA replication




‚Replication fork' in DNA replication

The mechanism of DNA replication is ‚semi-conservative'




Complementary replication is the simplest copying mechanism of RNA.
Complementarity is determined by Watson-Crick base pairs:

**G≡C** and **A=U**

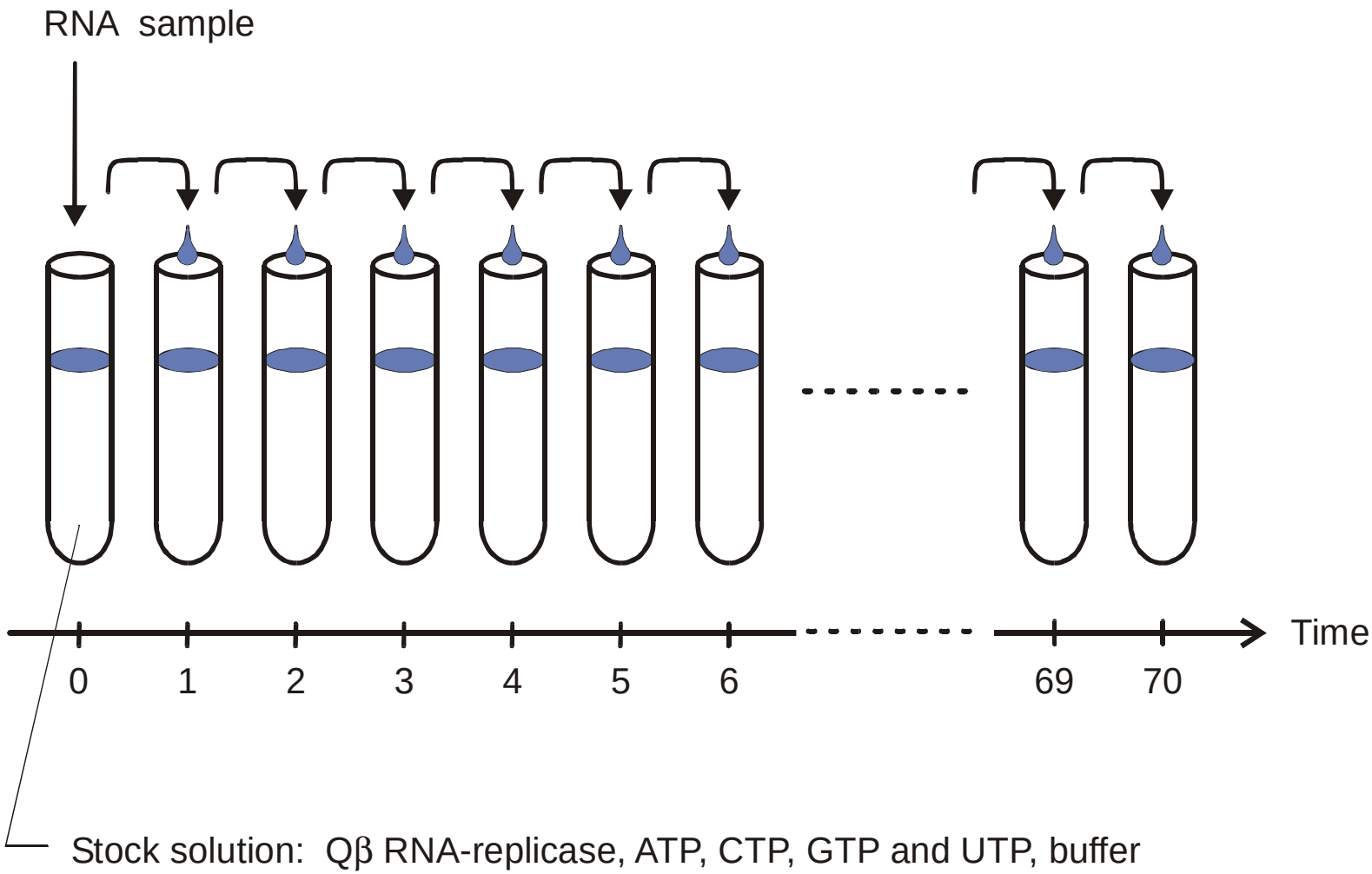
Drei notwendige Bedingungen für Darwinsche Evolution:

1. **Vermehrung,**
2. **Variation**, and
3. **Selektion**.

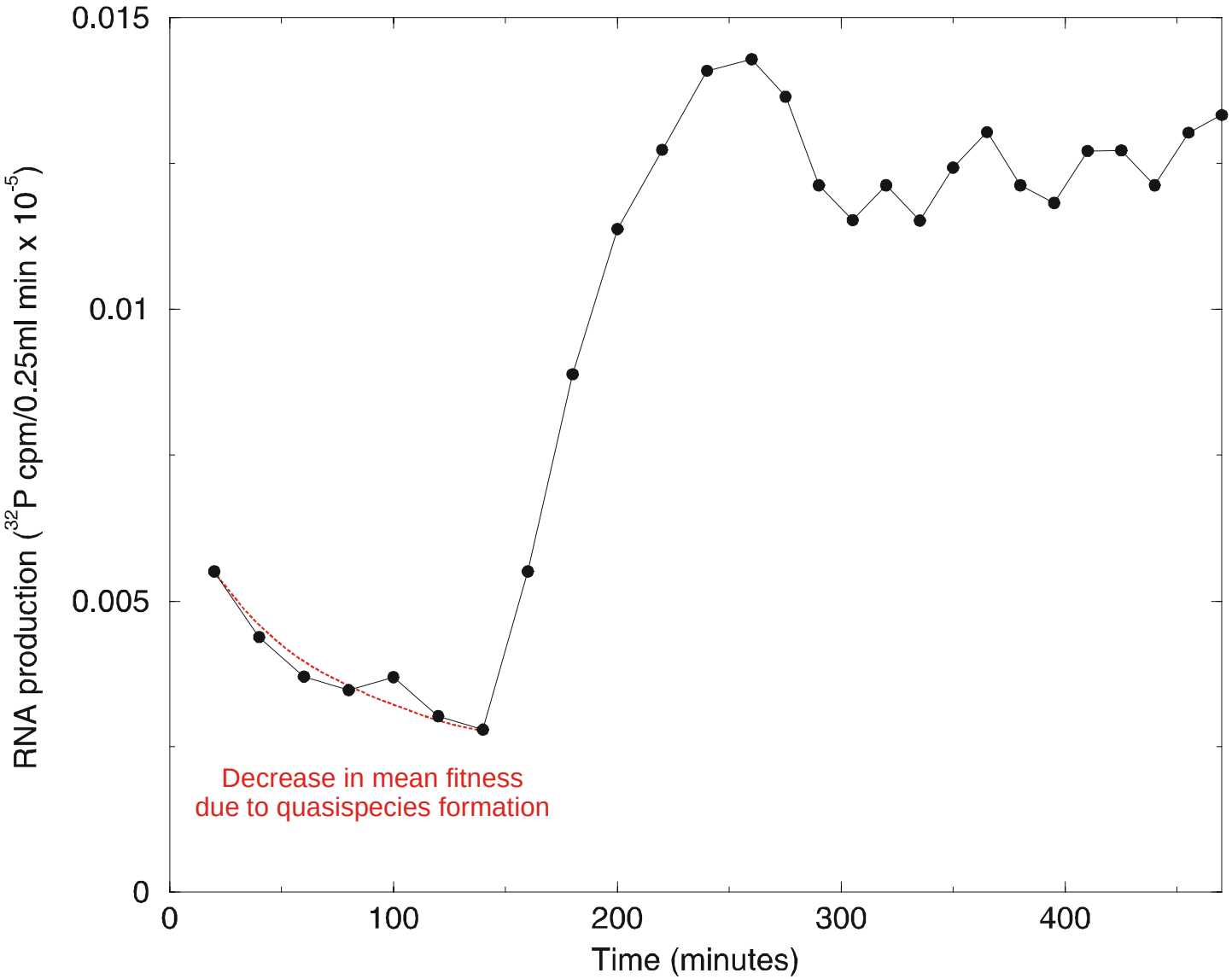
**Variation** in Form von Rekombination und/oder Mutation verändert die **Genotypen** wogegen **Selektion** nur auf den **Phänotypen** operiert.

Im Darwinschen Szenario treten **Variationen** in Form von Rekombinations- und/oder Mutationsereignissen **unkorreliert** mit ihren **Effekt auf den Selektionprocess** auf und erscheinen daher **zufällig**.

Alle drei Bedingungen werden nicht nur von zellulären Organismen erfüllt sondern auch von **Molekülen** in geigneten **zellfreien Assays**.

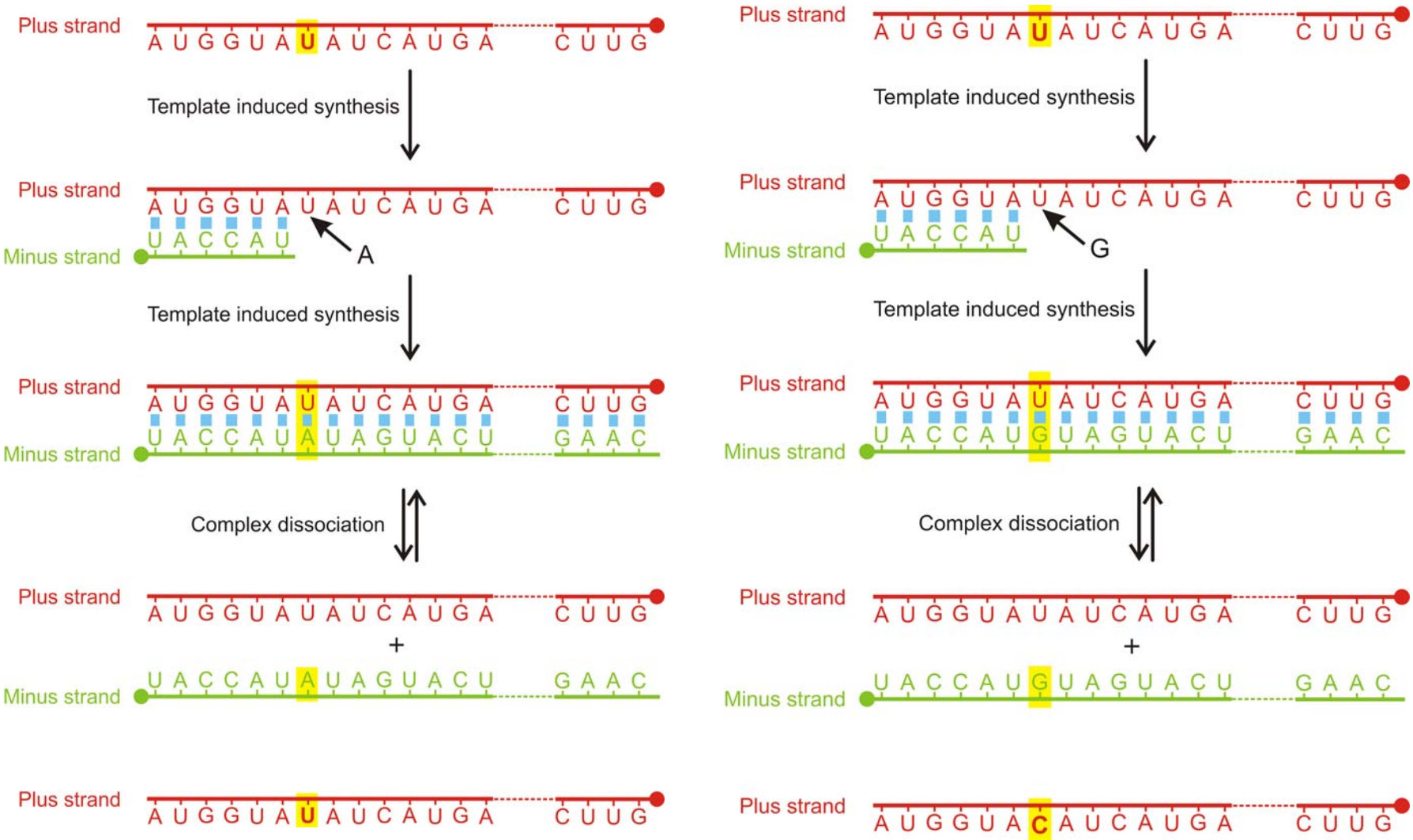



Application of the serial transfer technique to RNA evolution in the test tube

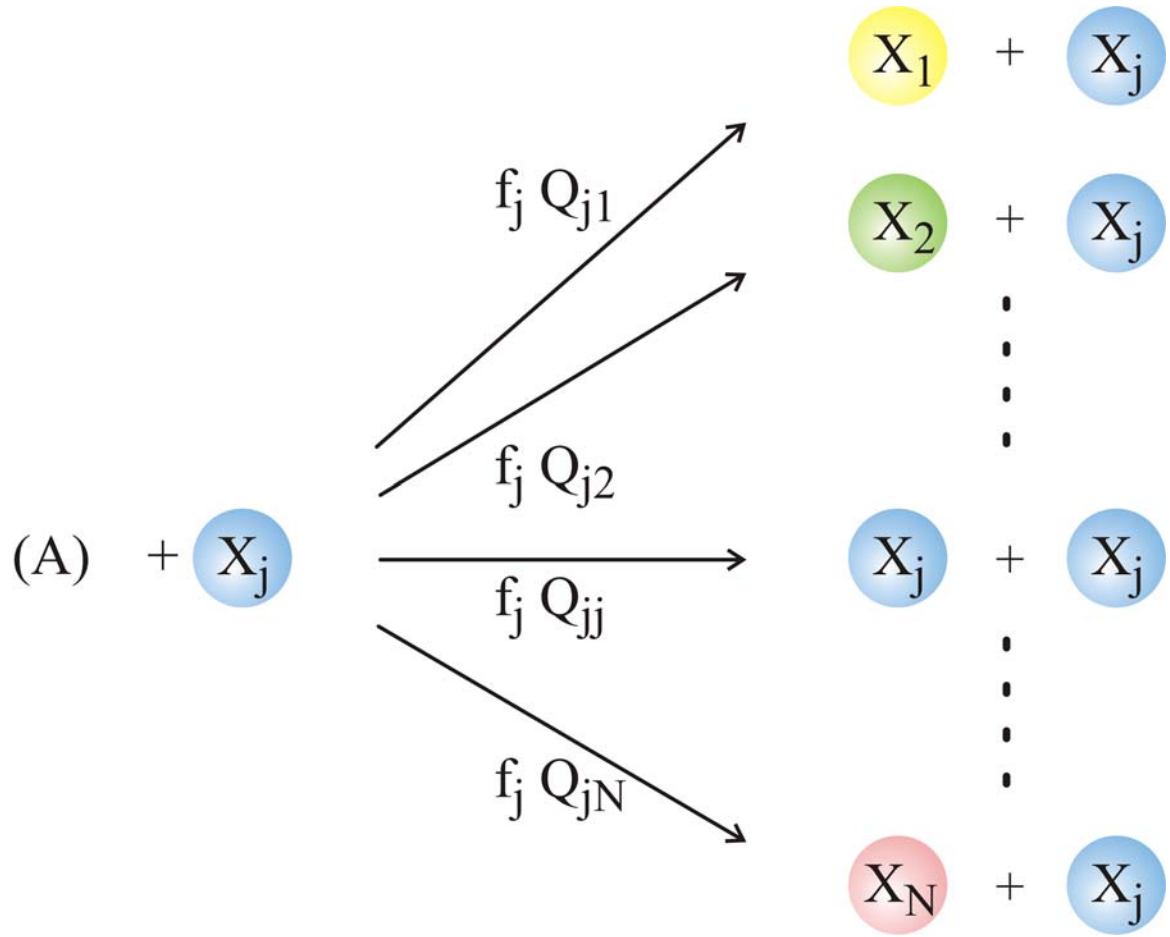



The increase in RNA production rate during a serial transfer experiment

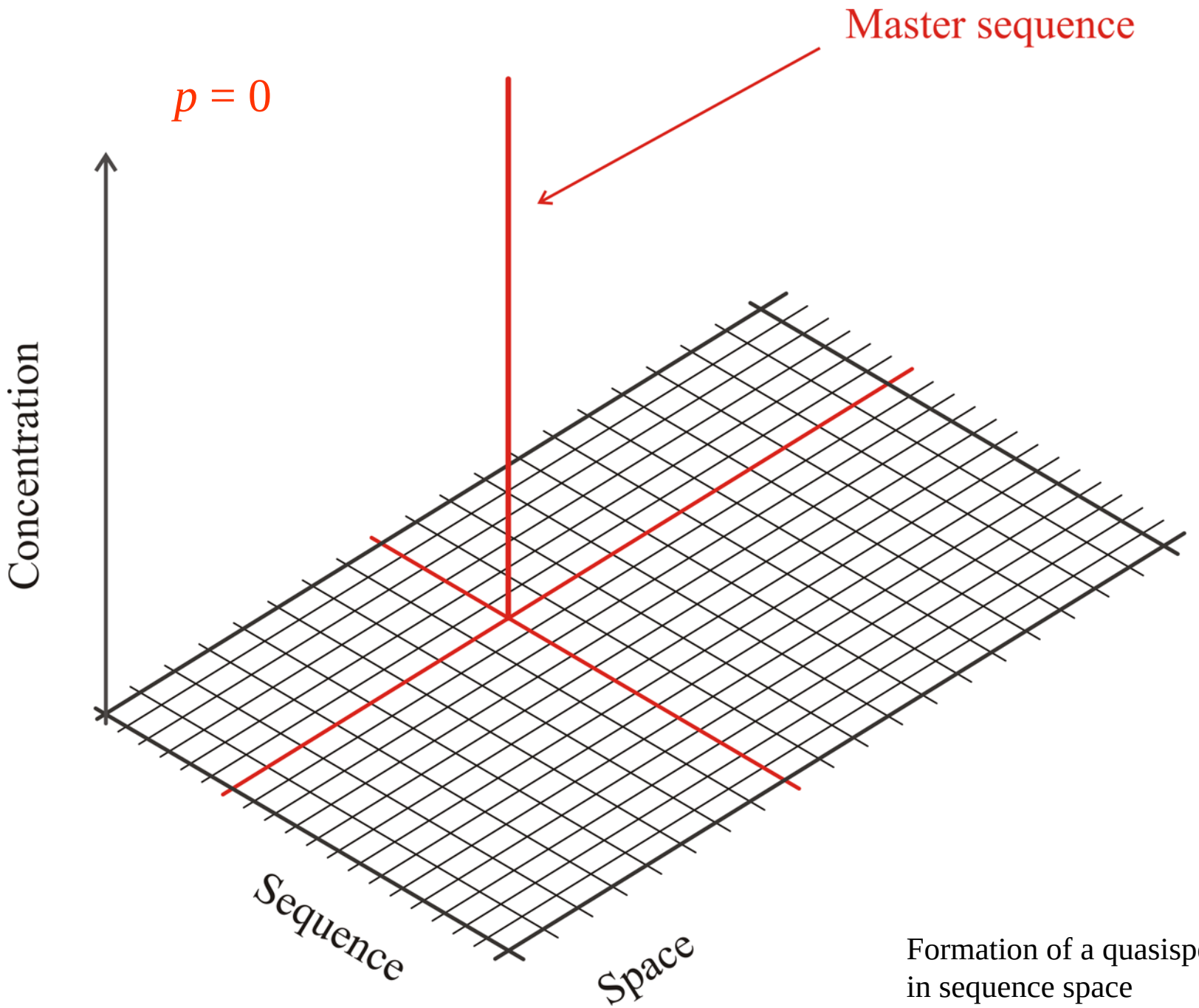
1. Musterbildung in Physik und Chemie
2. Muster in der Biologie
3. Darwins Prinzip der natürlichen Auslese
4. Vermehrung und Evolution von Molekülen
5. **Chemische Kinetik der molekularen Evolution**
6. Evolutionsexperimente mit Molekülen
7. Ursachen und Konsequenzen der Neutralität
8. Komplexität in der Biologie



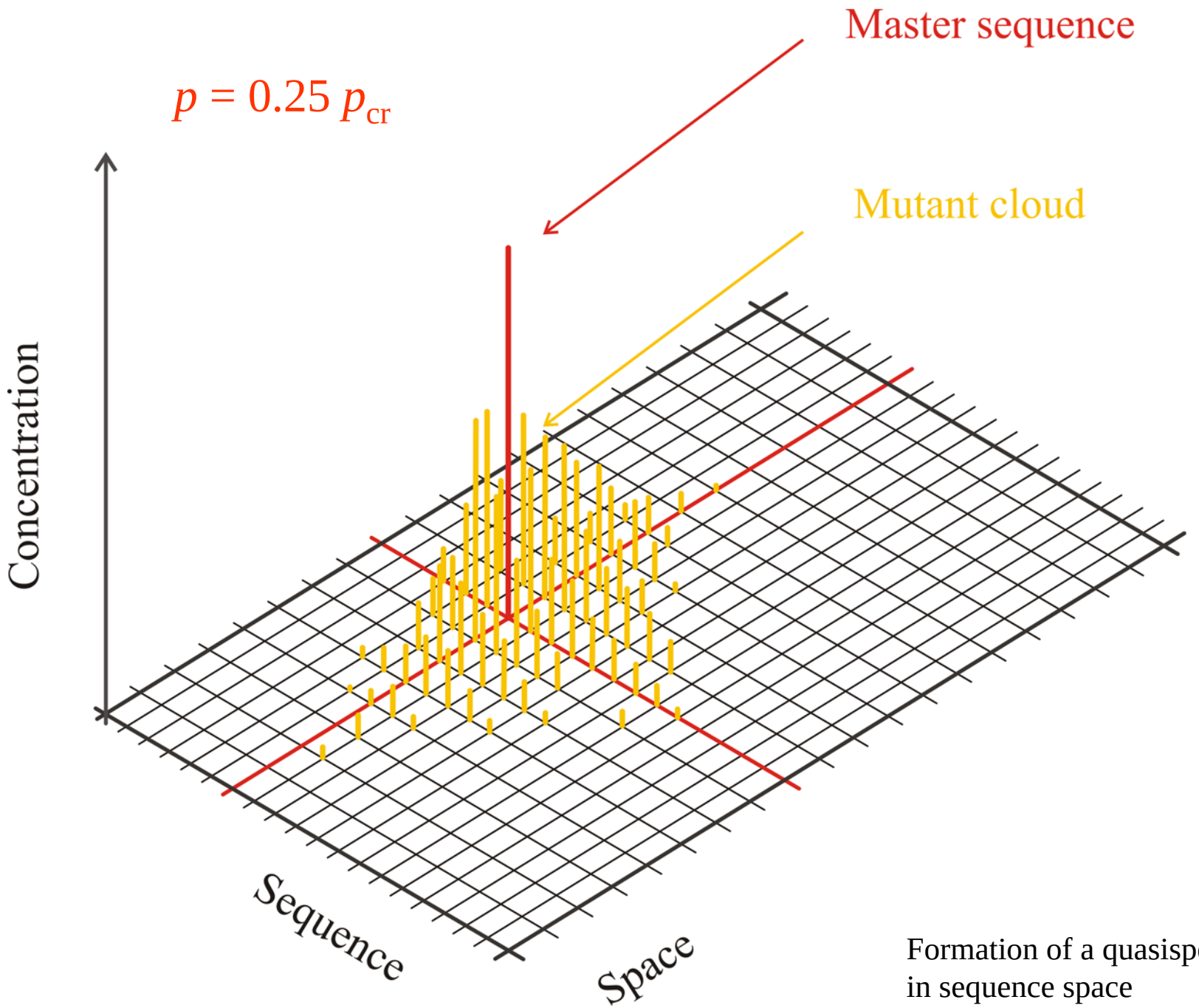

Replication and mutation are parallel chemical reactions.



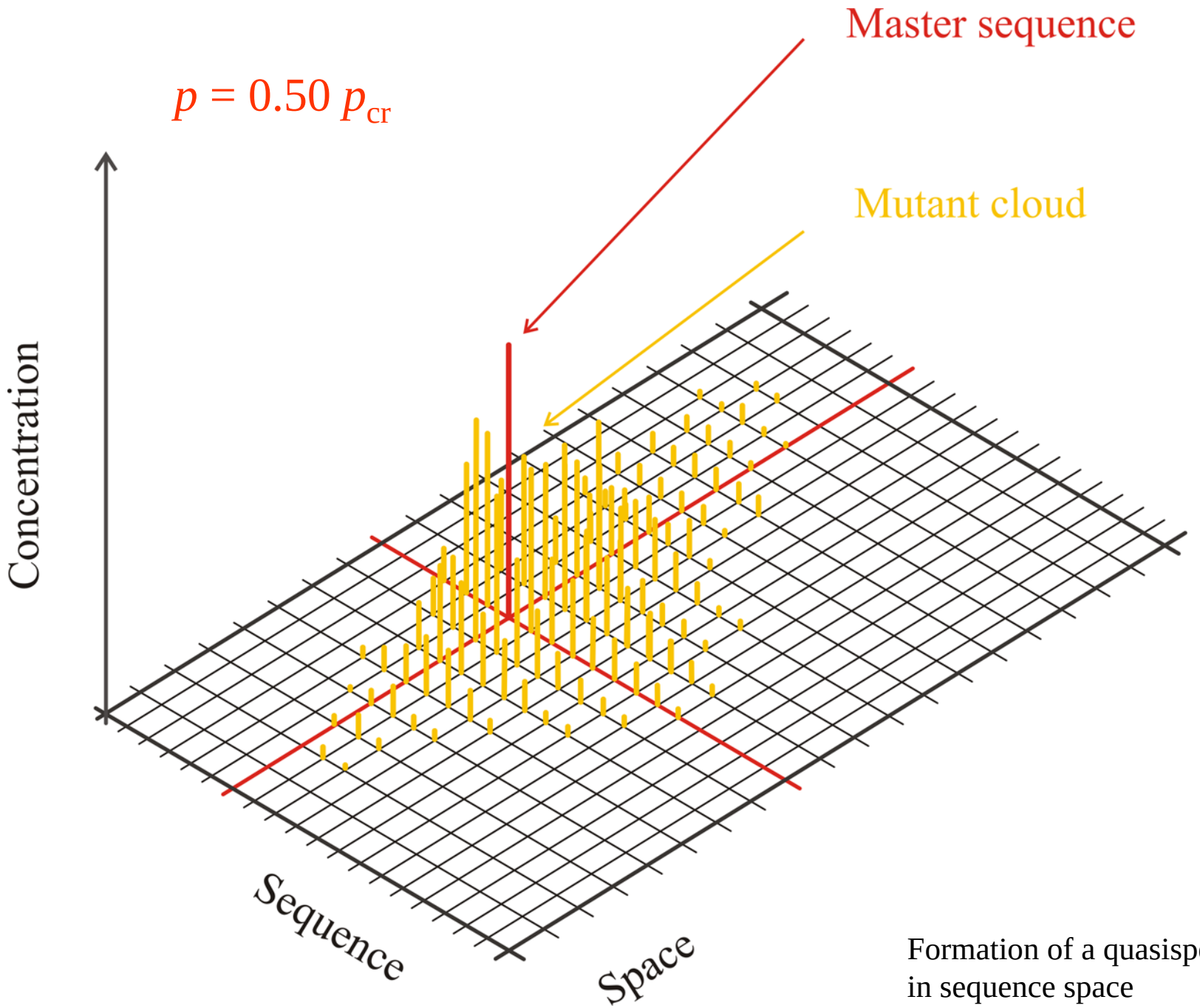

Chemical kinetics of replication and mutation as parallel reactions



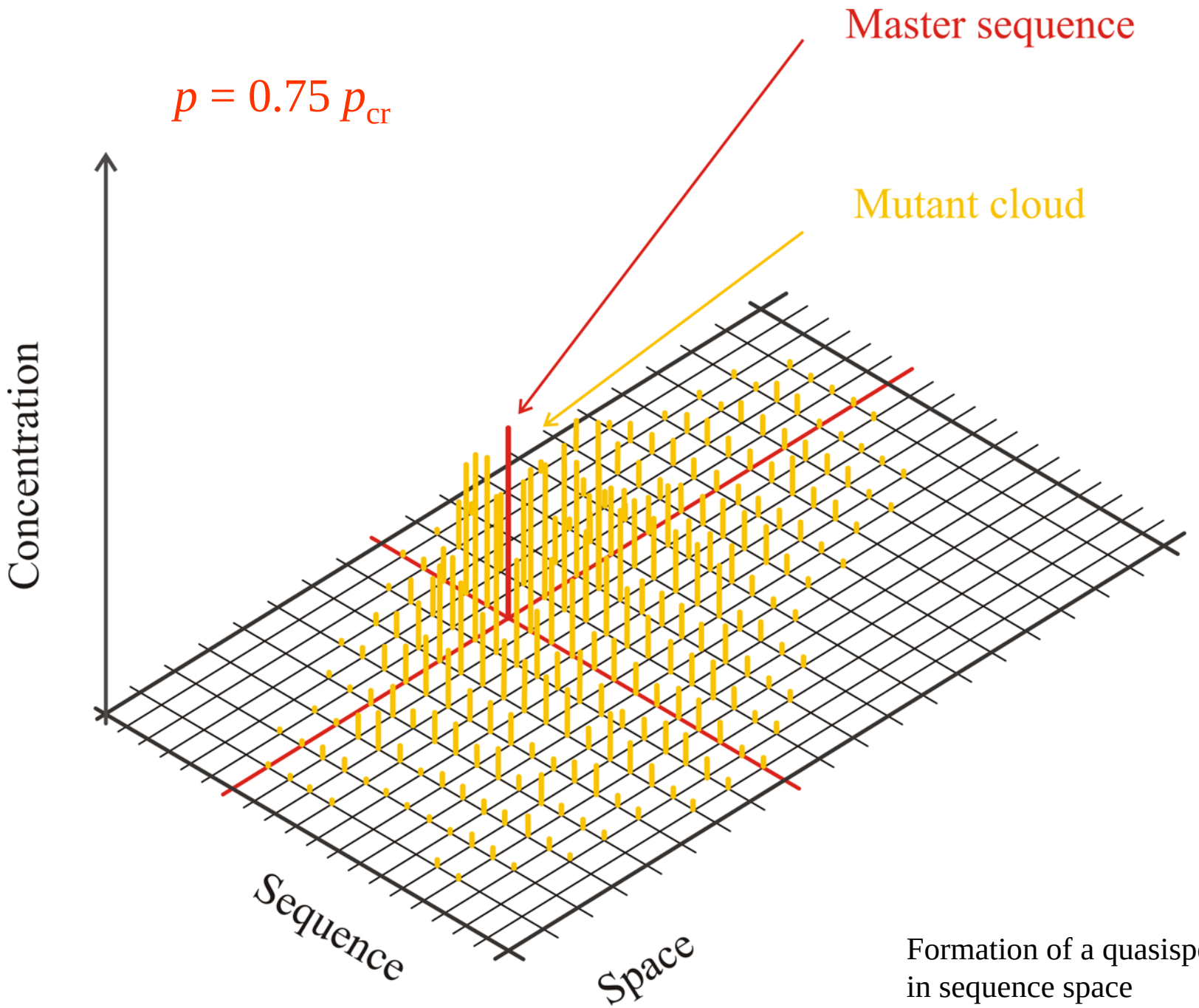

Formation of a quasispecies in sequence space




Formation of a quasispecies in sequence space

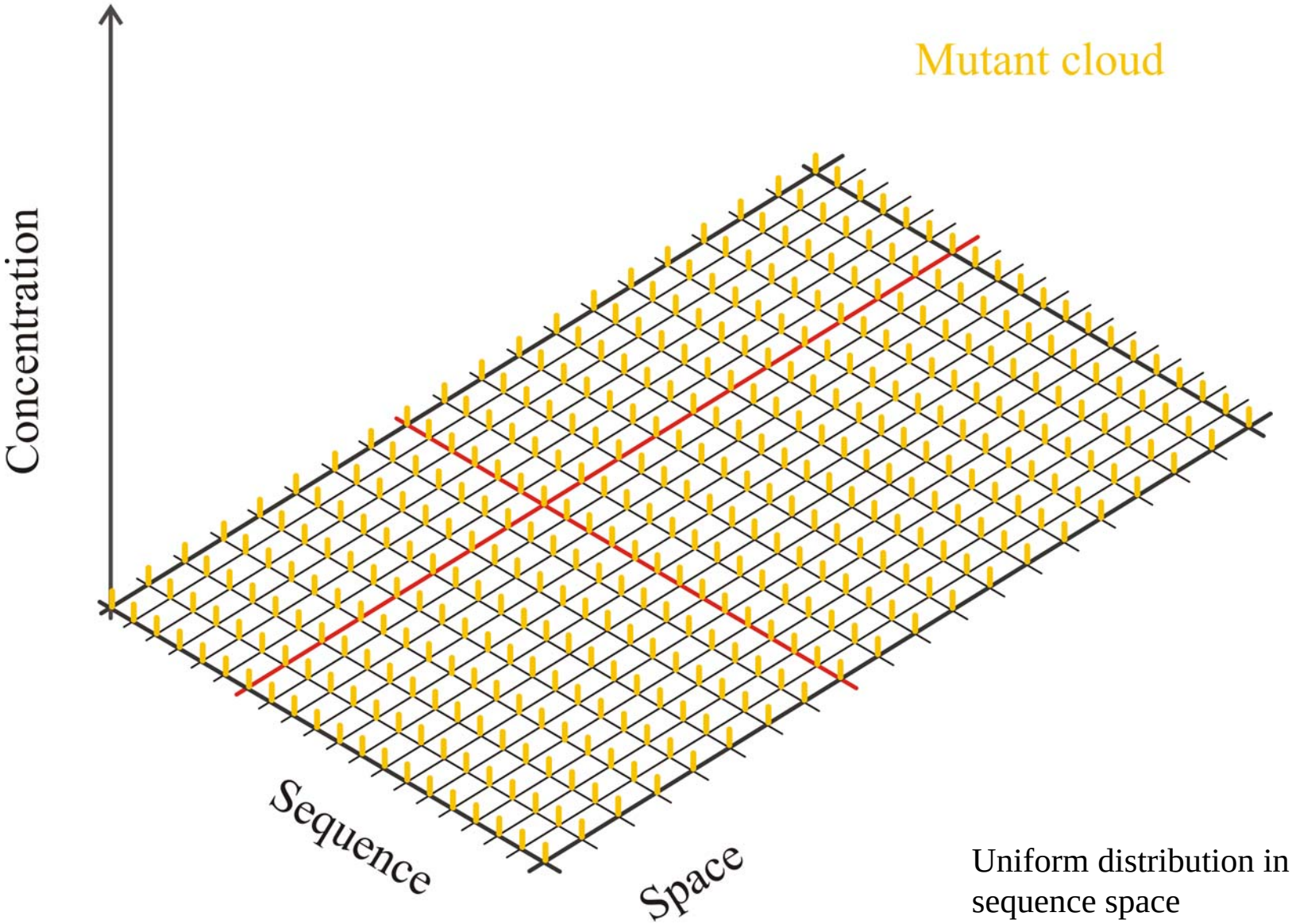



Formation of a quasispecies in sequence space

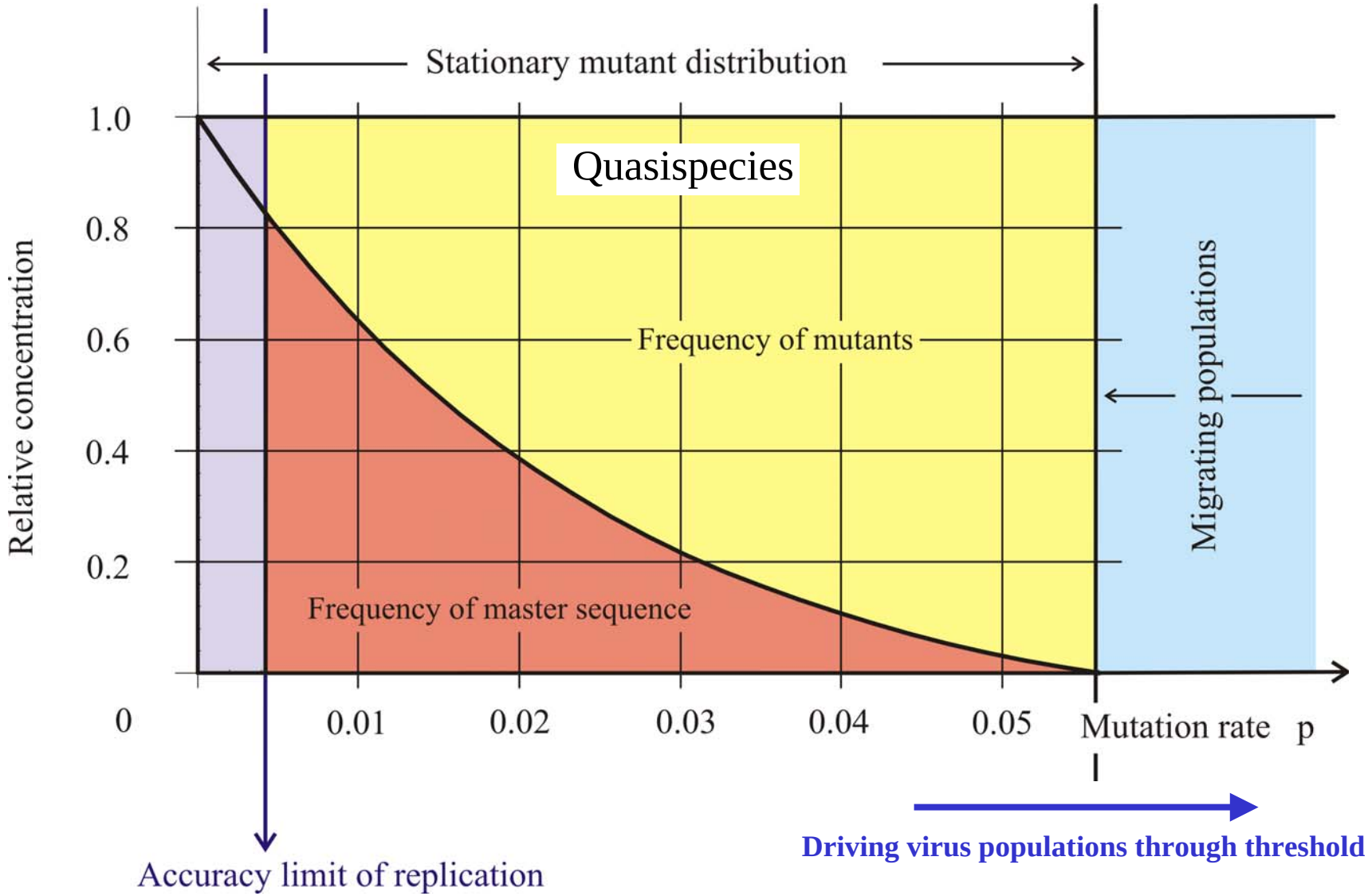



Formation of a quasispecies in sequence space

$p \geq p_{cr}$




Uniform distribution in sequence space




The error threshold in replication-mutation ensembles

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Virus Research 107 (2005) 115–116
Virus Research
www.elsevier.com/locate/virusres

Preface

# Antiviral strategy on the horizon

Error catastrophe had its conceptual origins in the middle of the XXth century, when the consequences of mutations on enzymes involved in protein synthesis, as a theory of aging. In those times biological processes were generally perceived differently from today. Infectious diseases were regarded as a fleeting nuisance which would be eliminated through the use of antibiotics and antiviral agents. Microbial variation, although known in some cases, was not thought to be a significant problem for disease control. Variation in differentiated organisms was seen as resulting essentially from exchanges of genetic material associated with sexual reproduction. The problem was to unveil the mechanisms of inheritance, expression of genetic information and metabolism. Few saw that genetic change is occurring at present in all organisms, and still fewer recognized Darwinian principles as essential to the biology of pathogenic viruses and cells. Population geneticists rarely used bacteria or viruses as experimental systems to define concepts in biological evolution. The extent of genetic polymorphism among individuals of the same biological species came as a surprise when the first results on comparison of electrophoretic mobility of enzymes were obtained. With the advent of in vitro DNA recombination, and rapid nucleic acid sequencing techniques, molecular analyses of genomes reinforced the conclusion of extreme inter-individual genetic variation within the same species. Now, due largely to spectacular progress in comparative genomics, we see cellular DNAs, both prokaryotic and eukaryotic, as highly dynamic. Most cellular processes, including such essential information-bearing and transferring events as genome replication, transcription and translation, are increasingly perceived as inherently inaccurate. Viruses, and in particular RNA viruses, are among the most extreme examples of exploitation of replication inaccuracy for survival.

Error catastrophe, or the loss of meaningful genetic information through excess genetic variation, was formulated in quantitative terms as a consequence of quasispecies theory, which was first developed to explain self-organization and adaptability of primitive replicons in early stages of life. Recently, a conceptual extension of error catastrophe that could be defined as "induced genetic deterioration" has emerged as a possible antiviral strategy. This is the topic of the current special issue of ***Virus Research***.

Few would nowadays doubt that one of the major obstacles for the control of viral disease is short-term adaptability of viral pathogens. Adaptability of viruses follows the same Darwinian principles that have shaped biological evolution over eons, that is, repeated rounds of reproduction with genetic variation, competition and selection, often perturbed by random events such as statistical fluctuations in population size. However, with viruses the consequences of the operation of these very same Darwinian principles are felt within very short times. Short-term evolution (within hours and days) can be also observed with some cellular pathogens, with subsets of normal cells, and cancer cells. The nature of RNA viral pathogens begs for alternative antiviral strategies, and forcing the virus to cross the critical error threshold for maintenance of genetic information is one of them.

The contributions to this volume have been chosen to reflect different lines of evidence (both theoretical and experimental) on which antiviral designs based on genetic deterioration inflicted upon viruses are being constructed. Theoretical studies have explored the copying fidelity conditions that must be fulfilled by any information-bearing replication system for the essential genetic information to be transmitted to progeny. Closely related to the theoretical developments have been numerous experimental studies on quasispecies dynamics and their multiple biological manifestations. The latter can be summarized by saying that RNA viruses, by virtue of existing as mutant spectra rather than defined genetic entities, remarkably expand their potential to overcome selective pressures intended to limit their replication. Indeed, the use of antiviral inhibitors in clinical practice and the design of vaccines for a number of major RNA virus-associated diseases, are currently presided by a sense of uncertainty. Another line of growing research is the enzymology of copying fidelity by viral replicases, aimed at understanding the molecular basis of mutagenic activities. Error catastrophe as a potential new antiviral strategy received an important impulse by the observation that ribavirin (a licensed antiviral nucleoside analogue) may be exerting, in some systems, its antiviral activity through enhanced mutagenesis. This has encouraged investigations on new mutagenic base analogues, some of them used in anticancer chemotherapy. Some chapters summarize these important biochemical studies on cell entry pathways and metabolism of mutagenic agents, that may find new applications as antiviral agents.

This volume intends to be basically a progress report, an introduction to a new avenue of research, and a realistic appraisal of the many issues that remain to be investigated. In this respect, I can envisage (not without many uncertainties) at least three lines of needed research: (i) One on further understanding of quasispecies dynamics in infected individuals to learn more on how to apply combinations of virus-specific mutagens and inhibitors in an effective way, finding synergistic combinations and avoiding antagonistic ones as well as severe clinical side effects. (ii) Another on a deeper understanding of the metabolism of mutagenic agents, in particular base and nucleoside analogues. This includes identification of the transporters that carry them into cells, an understanding of their metabolic processing, intracellular stability and alterations of nucleotide pools, among other issues. (iii) Still another line of needed research is the development of new mutagenic agents specific for viruses, showing no (or limited) toxicity for cells. Some advances may come from links with anticancer research, but others should result from the designs of new molecules, based on the structures of viral polymerases. I really hope that the reader finds this issue not only to be an interesting and useful review of the current situation in the field, but also a stimulating exposure to the major problems to be faced.

The idea to prepare this special issue came as a kind invitation of Ulrich Desselberger, former Editor of ***Virus Research***, and then taken enthusiastically by Luis Enjuanes, recently appointed as Editor of ***Virus Research***. I take this opportunity to thank Ulrich, Luis and the Editor-in-Chief of ***Virus Research***, Brian Mahy, for their continued interest and support to the research on virus evolution over the years.

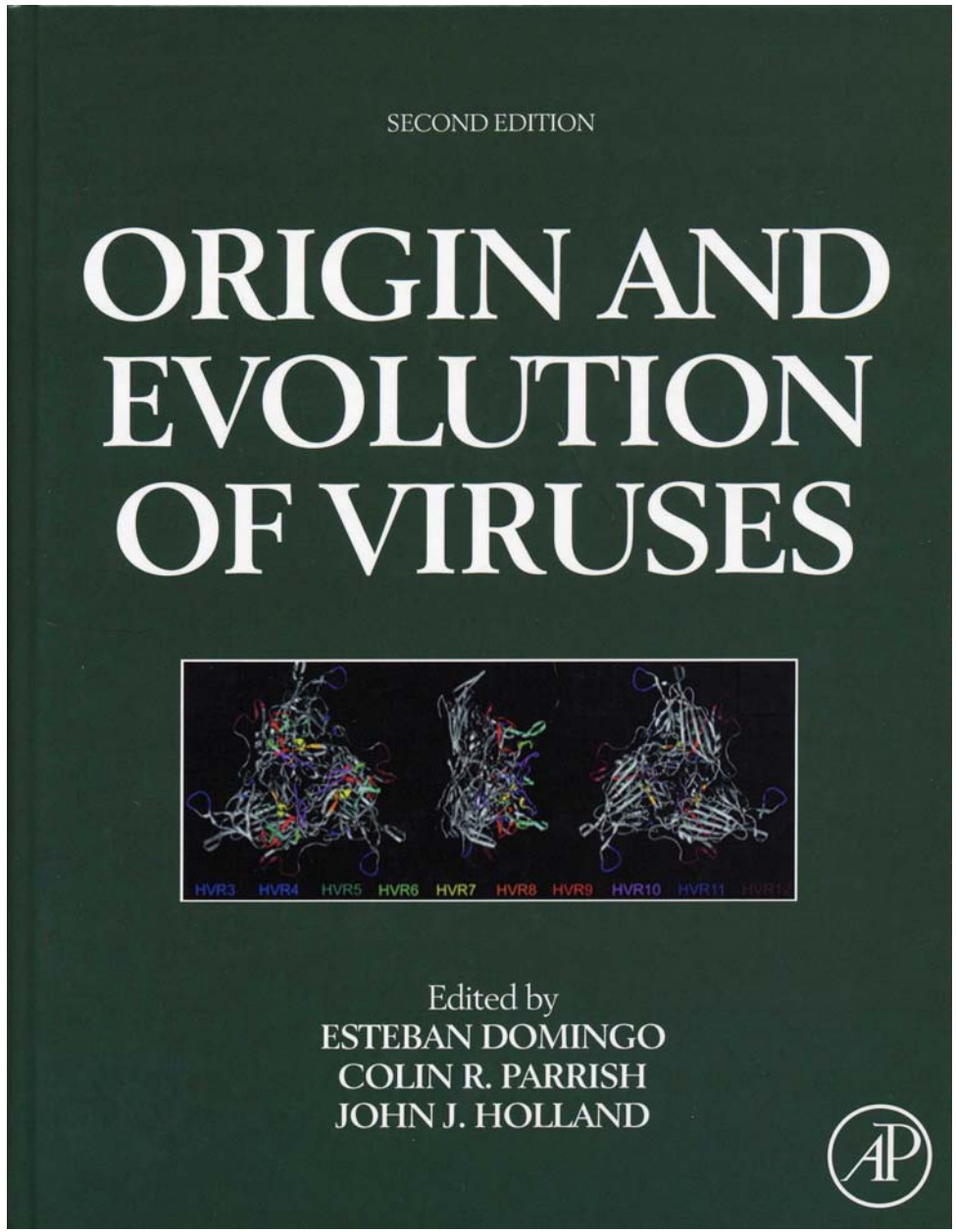
My thanks go also to the 19 authors who despite their busy schedules have taken time to prepare excellent manuscripts, to Elsevier staff for their prompt responses to my requests, and, last but not least, to Ms. Lucía Horrillo from Centro de Biología Molecular "Severo Ochoa" for her patient dealing with the correspondence with authors and the final organization of the issue.

Esteban Domingo
*Universidad Autónoma de Madrid*
*Centro de Biología Molecular "Severo Ochoa"*
*Consejo Superior de Investigaciones Científicas*
*Cantoblanco and Valdeolmos*
*Madrid, Spain*
Tel.: +34 91 497 84858/9; fax: +34 91 497 4799
*E-mail address:* edomingo@cbm.uam.es

Available online 8 December 2004

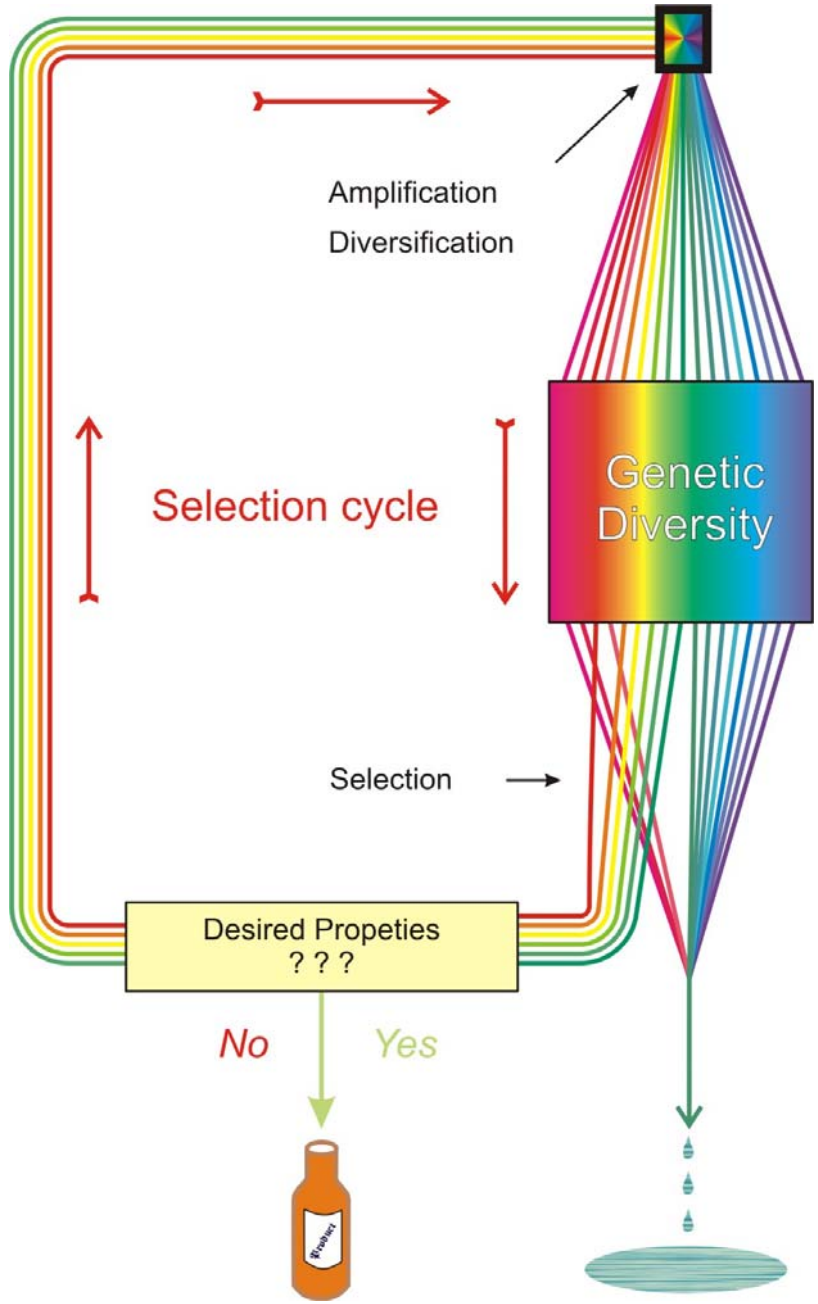

doi:10.1016/j.virusres.2004.11.001

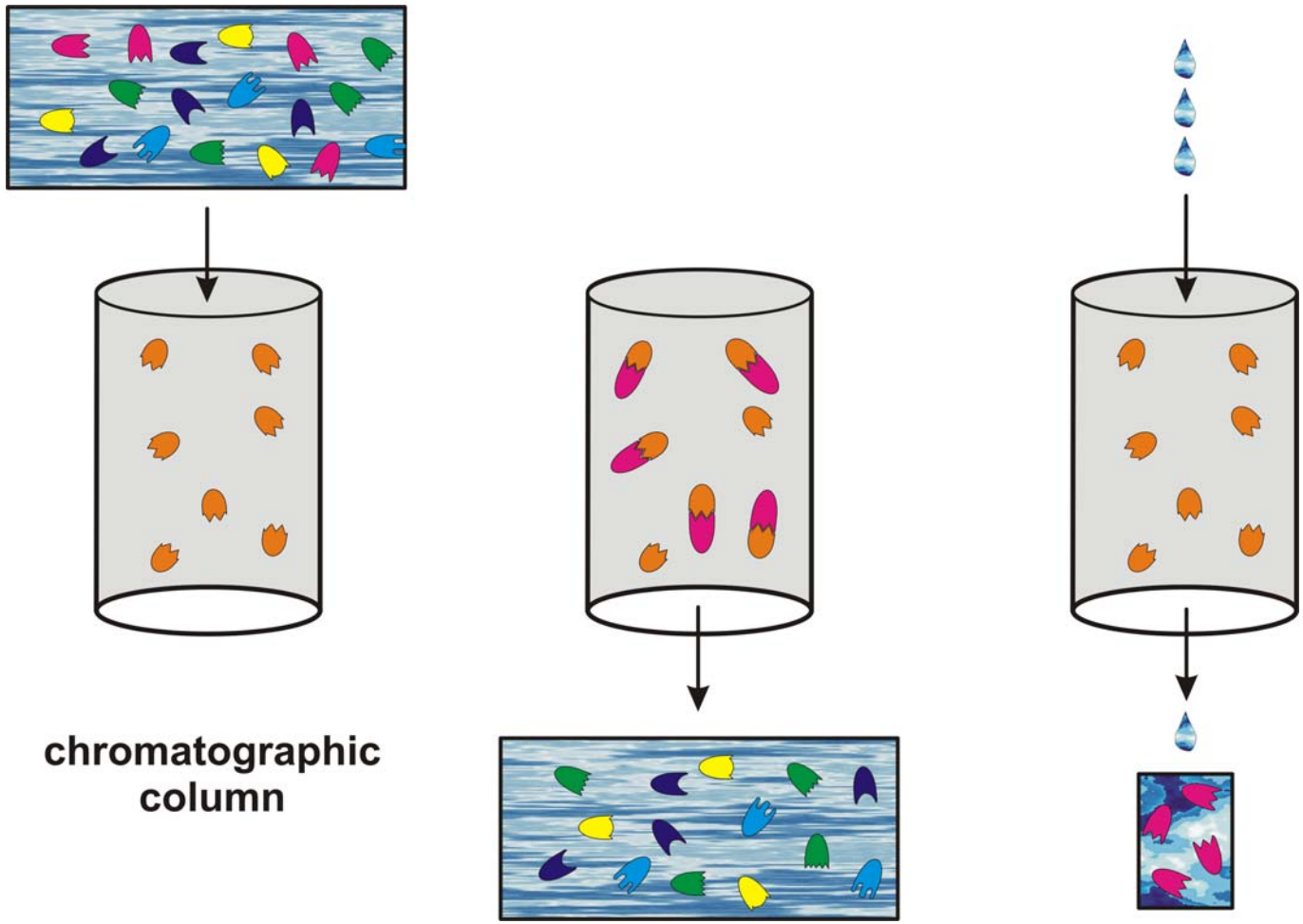
Molecular evolution of viruses



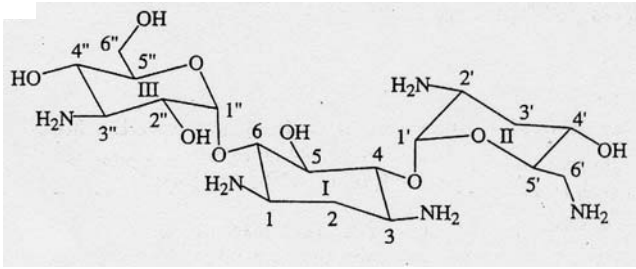
1. Musterbildung in Physik und Chemie
2. Muster in der Biologie
3. Darwins Prinzip der natürlichen Auslese
4. Vermehrung und Evolution von Molekülen
5. Chemische Kinetik der molekularen Evolution
6. **Evolutionsexperimente mit Molekülen**
7. Ursachen und Konsequenzen der Neutralität
8. Komplexität in der Biologie

An example of ‘artificial selection’ with RNA molecules or ‘breeding’ of biomolecules

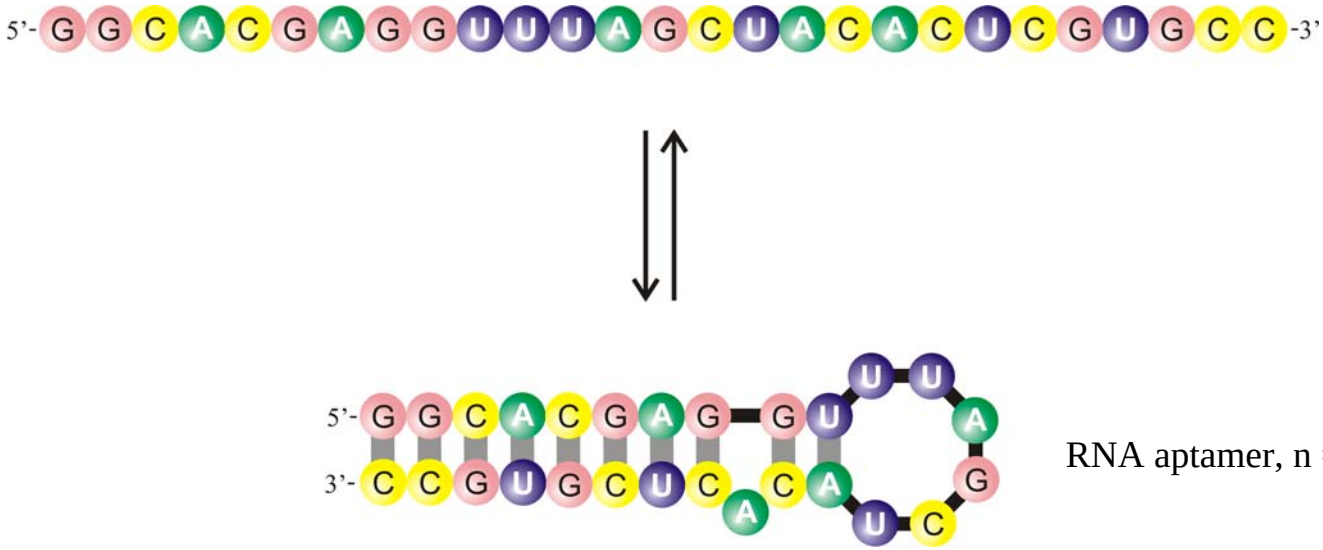





The SELEX-technique for evolutionary design of strongly binding molecules called aptamers



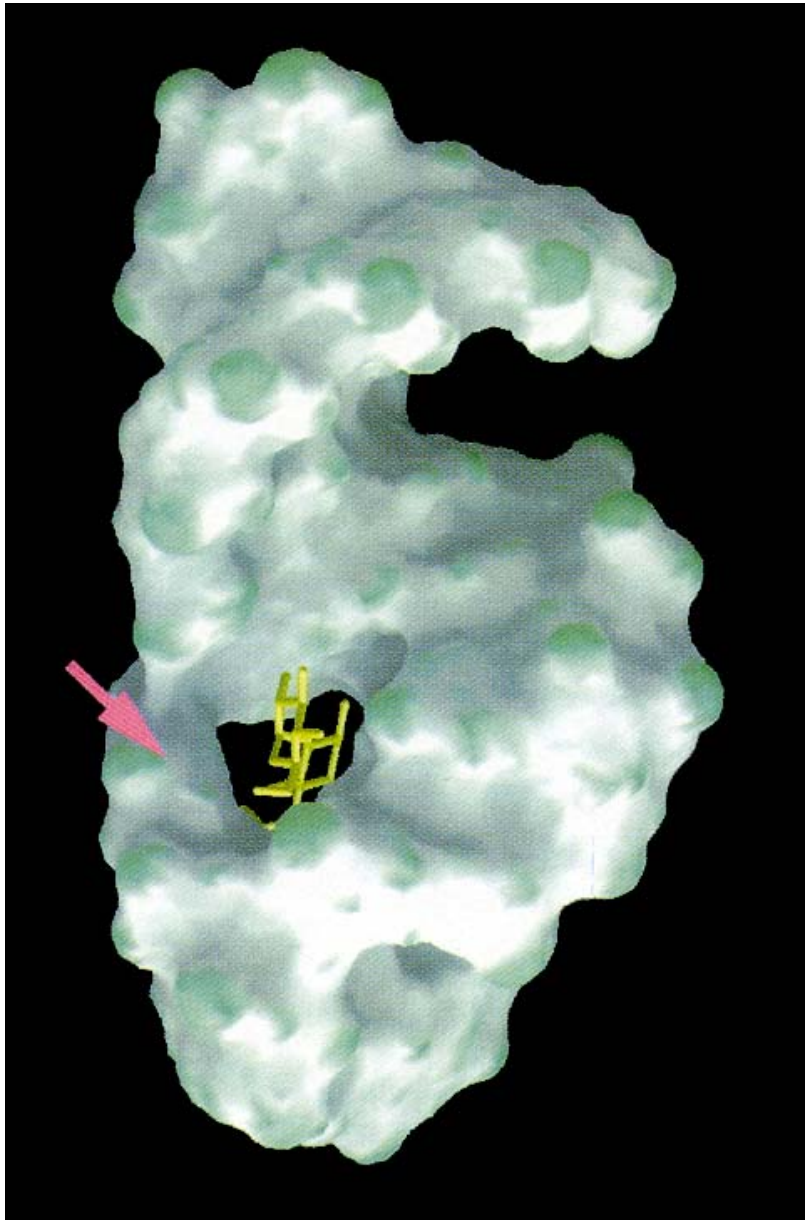

tobramycin




RNA aptamer, n = 27

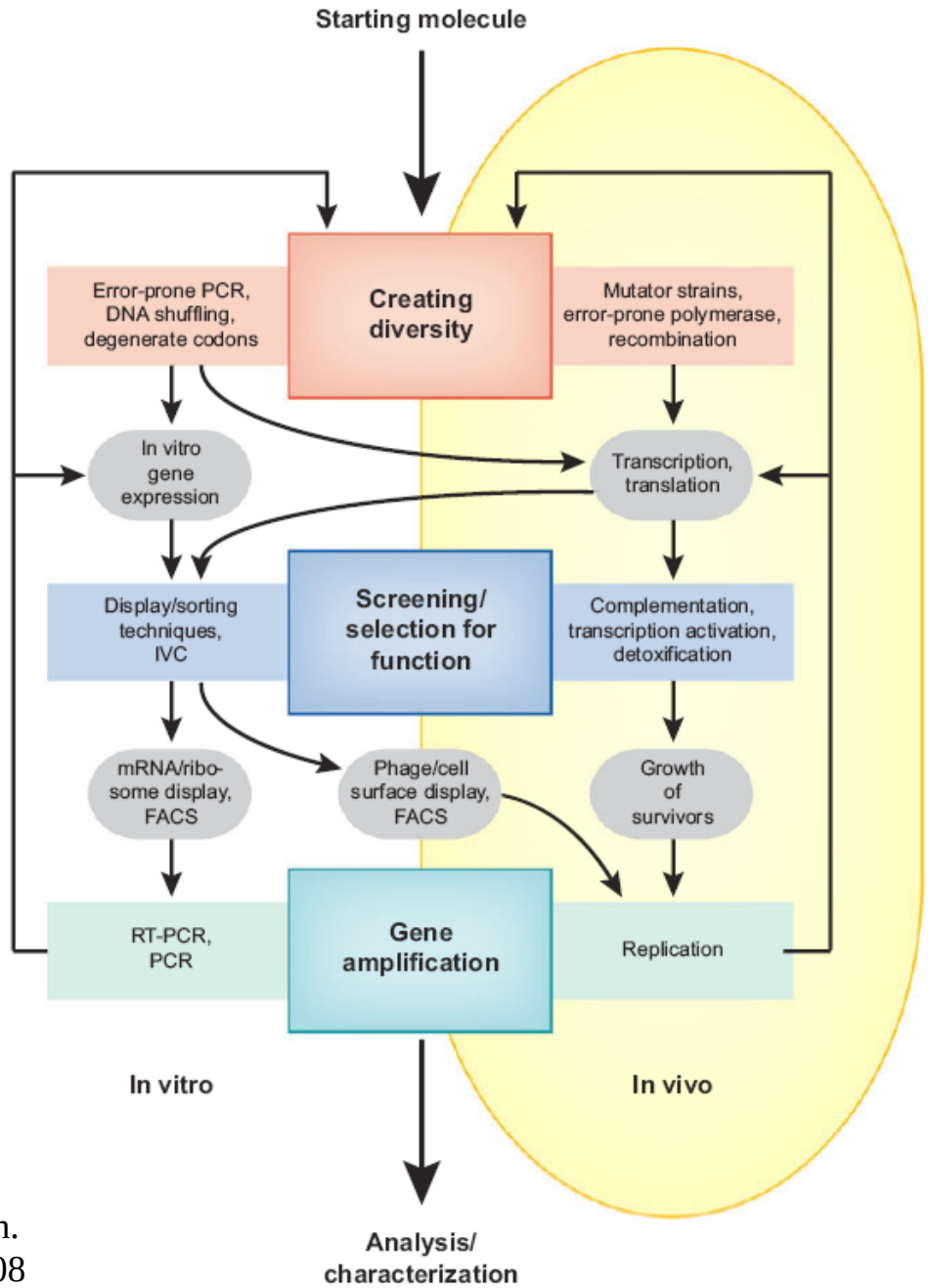
Formation of secondary structure of the tobramycin binding RNA aptamer with **$K_D$ = 9 nM**

L. Jiang, A. K. Suri, R. Fiala, D. J. Patel, ***Saccharide-RNA recognition in an aminoglycoside antibiotic-RNA aptamer complex.*** Chemistry & Biology **4**:35-50 (1997)



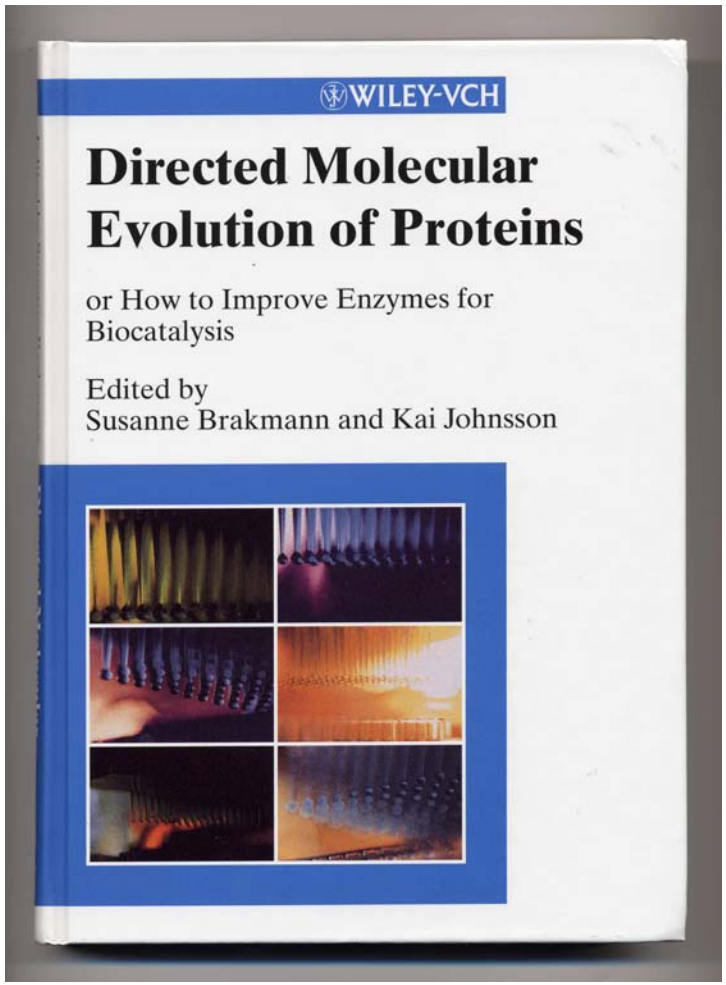
The three-dimensional structure of the tobramycin aptamer complex

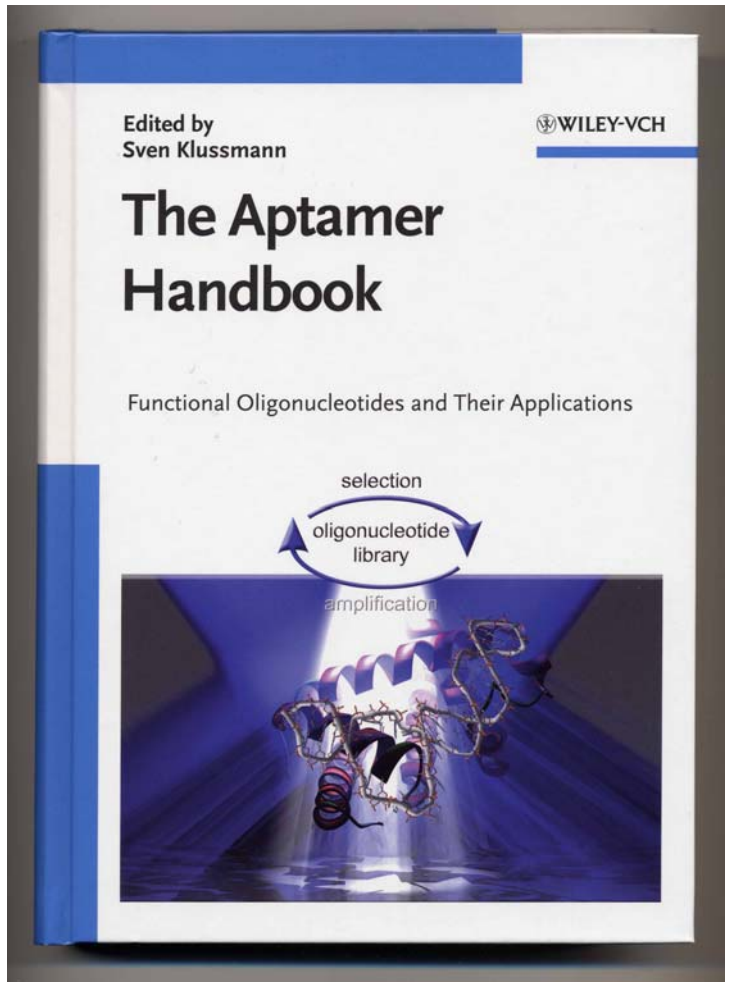
L. Jiang, A. K. Suri, R. Fiala, D. J. Patel, Chemistry & Biology **4**:35-50 (1997)




Schematic overview of the principal processes, strategies, and techniques of directed evolution. Today, numerous experimental methods are available to perform the fundamental processes of true Darwinian evolution (*central boxes*) in the laboratory, either in vivo within microorganisms or entirely in vitro in the test tube. Arrows indicate possible routes for connecting individual evolutionary steps. Abbreviations: PCR, polymerase chain reaction; RT-PCR, reverse transcription PCR; IVC, in vitro compartmentalization; FACS, fluorescence-activated cell sorting.

Christian Jäckel, Peter Kast, and Donald Hilvert.
Protein design by directed evolution.
*Annu.Rev.Biophys*. **37**:153-173, 2008




Application of molecular evolution to problems in biotechnology

**Artificial evolution in biotechnology and pharmacology**

G.F. Joyce. 2004. Directed evolution of nucleic acid enzymes. *Annu.Rev.Biochem*. **73**:791-836.

C. Jäckel, P. Kast, and D. Hilvert. 2008. Protein design by directed evolution. *Annu.Rev.Biophys*. **37**:153-173.

S.J. Wrenn and P.B. Harbury. 2007. Chemical evolution as a tool for molecular discovery. *Annu.Rev.Biochem*. **76**:331-349.

1. Musterbildung in Physik und Chemie
2. Muster in der Biologie
3. Darwins Prinzip der natürlichen Auslese
4. Vermehrung und Evolution von Molekülen
5. Chemische Kinetik der molekularen Evolution
6. Evolutionsexperimente mit Molekülen
7. **Ursachen und Konsequenzen der Neutralität**
8. Komplexität in der Biologie

# Was bedeutet Neutralität ?

Selektive Neutralität =
= mehrere Genotypen weisen identische Fitness auf.

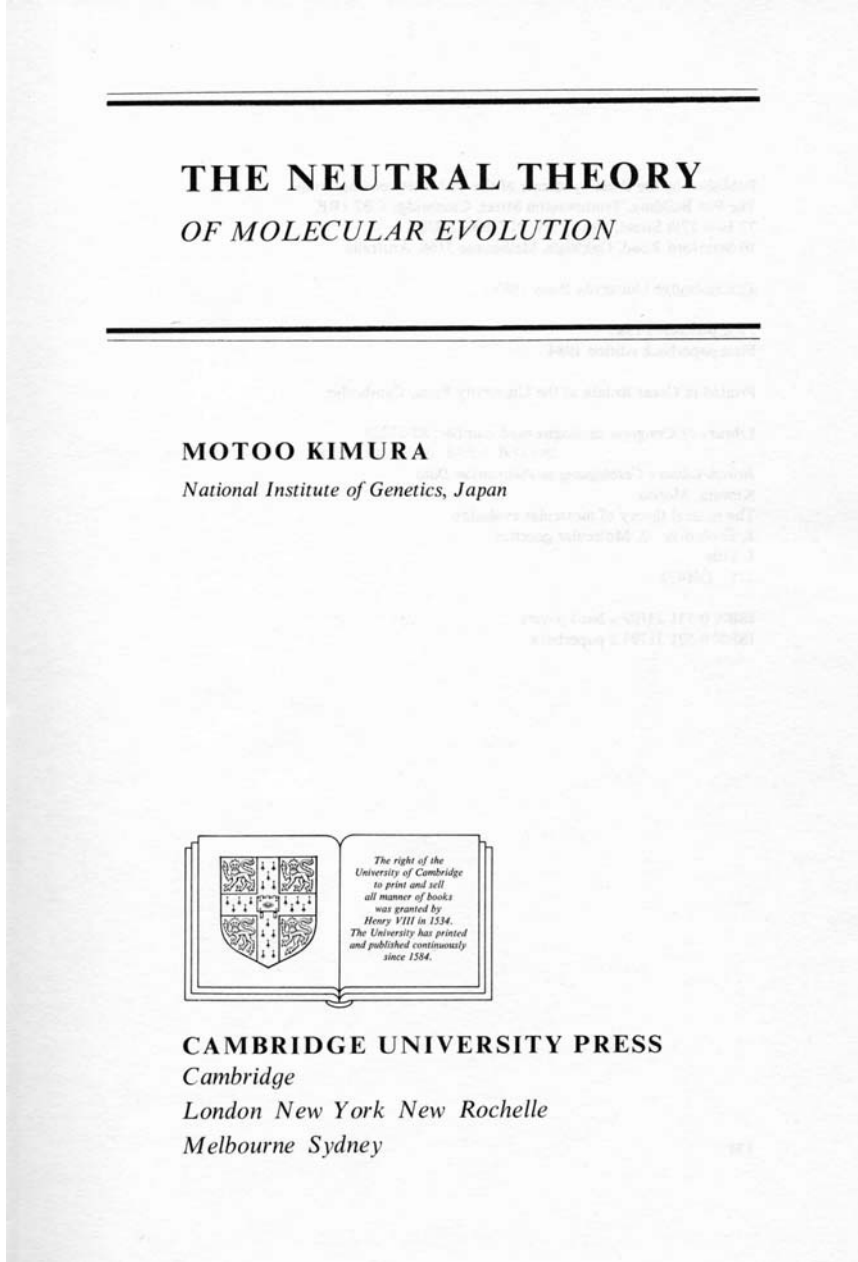
Strukturelle Neutralität =
= mehrere Genotypen bilden identische Strukturen aus.

MOTOO KIMURA

Motoo Kimuras Populationsgenetik der neutralen Evolution.

Evolutionary rate at the molecular level. *Nature* **217**: 624-626, 1955.

*The Neutral Theory of Molecular Evolution*. Cambridge University Press. Cambridge, UK, 1983.



THE NEUTRAL THEORY

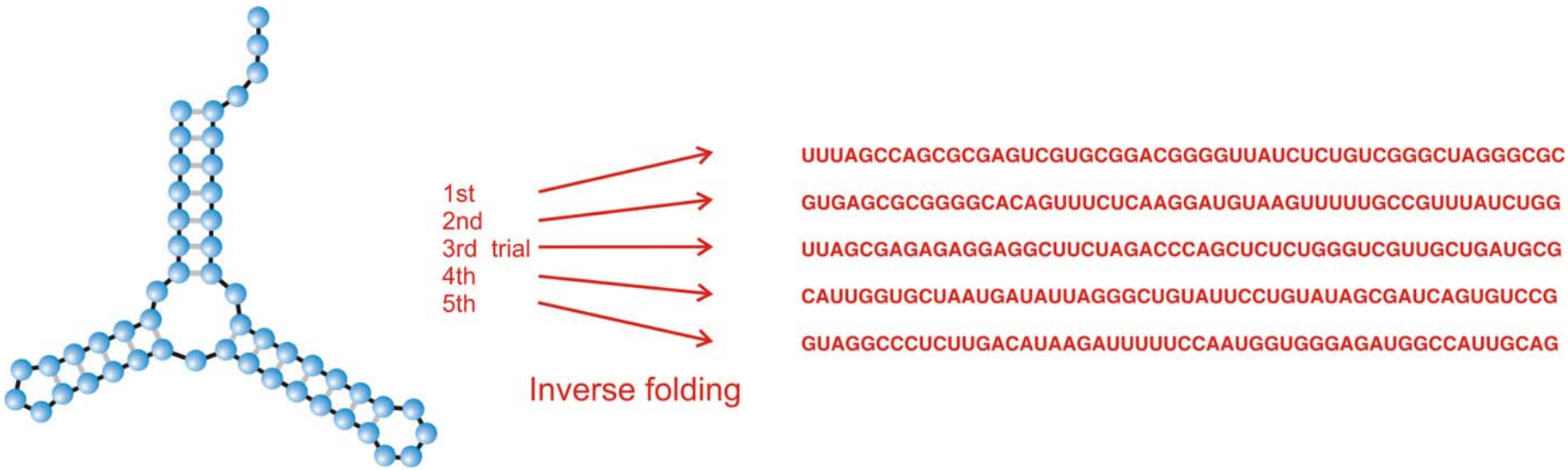
*OF MOLECULAR EVOLUTION*

MOTOO KIMURA

*National Institute of Genetics, Japan*



CAMBRIDGE UNIVERSITY PRESS
*Cambridge*
*London New York New Rochelle*
*Melbourne Sydney*




The inverse folding algorithm searches for sequences that form a given RNA structure.

# Evolution *in silico*

W. Fontana, P. Schuster,
*Science* **280** (1998), 1451-1455



random individuals. The primer pair used for genomic DNA amplification is 5′-TCTCCCTGGATTCT-CATTTA-3′ (forward) and 5′-TCTTTGTCTTCTGT-TCCACC-3′ (reverse). Reactions were performed in 25 μl using 1 unit of Taq DNA polymerase with each primer at 0.4 μM; 200 μM each dATP, dTTP, dGTP, and dCTP; and PCR buffer [10 mM tris-HCl (pH 8.3), 50 mM $KCl_2$,1.5 mM $MgCl_2$] in a cycle condition of 94°C for 1 min and then 35 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 30 s followed by 72°C for 6 min. PCR products were purified (Qiagen), digested with Xmn I, and separated in a 2% agarose gel.

32. A nonsense mutation may affect mRNA stability and result in degradation of the transcript [L. Maquat, *Am. J. Hum. Genet.* **59**, 279 (1996)].
33. Data not shown; a dot blot with poly (A)$^+$ RNA from 50 human tissues (The Human RNA Master Blot, 7770-1, Clontech Laboratories) was hybridized with a probe from exons 29 to 47 of *MYO15* using the same condition as Northern blot analysis (*13*).
34. Smith–Magenis syndrome (SMS) is due to deletions of 17p11.2 of various sizes, the smallest of which includes *MYO15* and perhaps 20 other genes [(*6*); K-S Chen, L. Potocki, J. R. Lupski, *MRDD Res. Rev.* **2**, 122 (1996)]. *MYO15* expression is easily detected in the pituitary gland (data not shown). Haploinsufficiency for *MYO15* may explain a portion of the SMS phenotype such as short stature. Moreover, a few SMS patients have sensorineural hearing loss, possibly because of a point mutation in *MYO15* in trans to the SMS 17p11.2 deletion.
35. R. A. Fridell, data not shown.
36. K. B. Avraham *et al.*, *Nature Genet.* **11**, 369 (1995); X-Z. Liu *et al.*, *ibid.* **17**, 268 (1997); F. Gibson *et al.*, *Nature* **374**, 62 (1995); D. Weil *et al.*, *ibid.*, p. 60.
37. RNA was extracted from cochlea (membranous labyrinths) obtained from human fetuses at 18 to 22 weeks of development in accordance with guidelines established by the Human Research Committee at the Brigham and Women's Hospital. Only samples without evidence of degradation were pooled for poly (A)$^+$ selection over oligo(dT) columns. First-strand cDNA was prepared using an Advantage RT-for-PCR kit (Clontech Laboratories). A portion of the first-strand cDNA (4%) was amplified by PCR with Advantage cDNA polymerase mix (Clontech Laboratories) using human *MYO15*-specific oligonucleotide primers (forward, 5′-GCATGACCTGCCGGCTAAT-GGG-3′; reverse, 5′-CTCACGGCTTCTGCATGGT-GCTCGGCTGGC-3′). Cycling conditions were 40 s at 94°C; 40 s at 66°C (3 cycles), 60°C (5 cycles), and 55°C (29 cycles); and 45 s at 68°C. PCR products were visualized by ethidium bromide staining after fractionation in a 1% agarose gel. A 688-bp PCR product is expected from amplification of the human *MYO15* cDNA. Amplification of human genomic DNA with this primer pair would result in a 2903-bp fragment.

38. We are grateful to the people of Bengkala, Bali, and the two families from India. We thank J. R. Lupski and K.-S. Chen for providing the human chromosome 17 cosmid library. For technical and computational assistance, we thank N. Dietrich, M. Fergusson, A. Gupta, E. Sorbello, R. Torkzadeh, C. Varner, M. Walker, G. Bouffard, and S. Beckstrom-Sternberg (National Institutes of Health Intramural Sequencing Center). We thank J. T. Hinnant, I. N. Arhya, and S. Winata for assistance in Bali, and T. Barber, S. Sullivan, E. Green, D. Drayna, and J. Battey for helpful comments on this manuscript. Supported by the National Institute on Deafness and Other Communication Disorders (NIDCD) (Z01 DC 00035-01 and Z01 DC 00038-01 to T.B.F. and E.R.W. and R01 DC 03402 to C.C.M.), the National Institute of Child Health and Human Development (R01 HD30428 to S.A.C.) and a National Science Foundation Graduate Research Fellowship to F.J.P. This paper is dedicated to J. B. Snow Jr. on his retirement as the Director of the NIDCD.

9 March 1998; accepted 17 April 1998

## Continuity in Evolution: On the Nature of Transitions

Walter Fontana and Peter Schuster

To distinguish continuous from discontinuous evolutionary change, a relation of nearness between phenotypes is needed. Such a relation is based on the probability of one phenotype being accessible from another through changes in the genotype. This nearness relation is exemplified by calculating the shape neighborhood of a transfer RNA secondary structure and provides a characterization of discontinuous shape transformations in RNA. The simulation of replicating and mutating RNA populations under selection shows that sudden adaptive progress coincides mostly, but not always, with discontinuous shape transformations. The nature of these transformations illuminates the key role of neutral genetic drift in their realization.

A much-debated issue in evolutionary biology concerns the extent to which the history of life has proceeded gradually or has been punctuated by discontinuous transitions at the level of phenotypes (*1*). Our goal is to make the notion of a discontinuous transition more precise and to understand how it arises in a model of evolutionary adaptation.

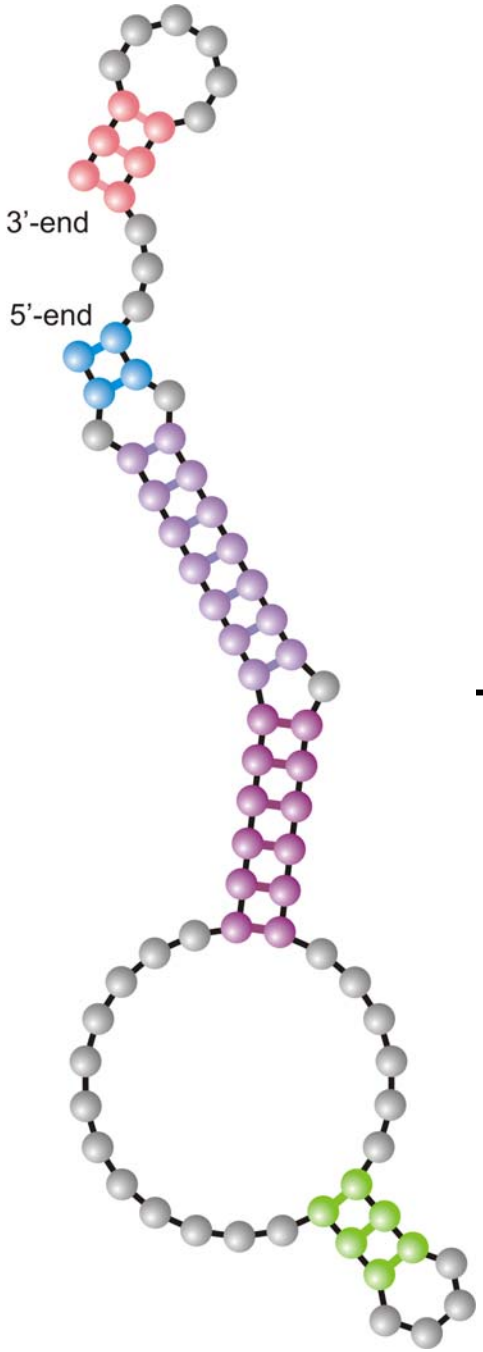
We focus on the narrow domain of RNA secondary structure, which is currently the simplest computationally tractable, yet realistic phenotype (*2*). This choice enables the definition and exploration of concepts that may prove useful in a wider context. RNA secondary structures represent a coarse level of analysis compared with the three-dimensional structure at atomic resolution. Yet, secondary structures are empirically well defined and obtain their biophysical and biochemical importance from being a scaffold for the tertiary structure. For the sake of brevity, we shall refer to secondary structures as "shapes." RNA combines in a single molecule both genotype (replicatable sequence) and phenotype (selectable shape), making it ideally suited for in vitro evolution experiments (*3*, *4*).

To generate evolutionary histories, we used a stochastic continuous time model of an RNA population replicating and mutating in a capacity-constrained flow reactor under selection (*5*, *6*). In the laboratory, a goal might be to find an RNA aptamer binding specifically to a molecule (*4*). Although in the experiment the evolutionary end product was unknown, we thought of its shape as being specified implicitly by the imposed selection criterion. Because our intent is to study evolutionary histories rather than end products, we defined a target shape in advance and assumed the replication rate of a sequence to be a function of the similarity between its shape and the target. An actual situation may involve more than one best shape, but this does not affect our conclusions.

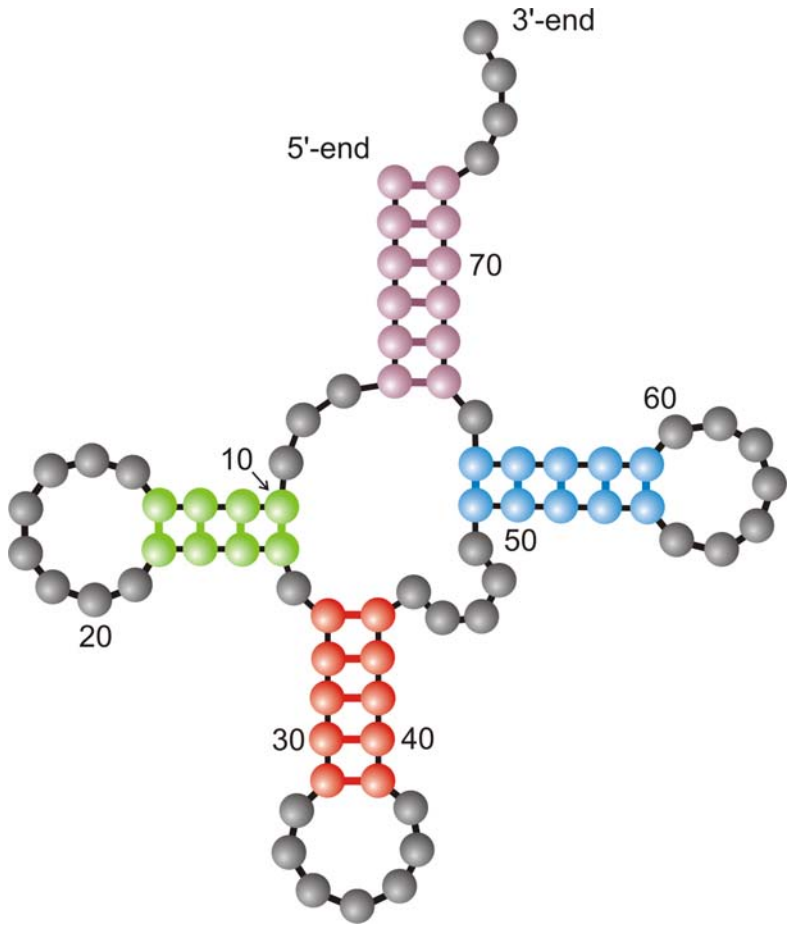
An instance representing in its qualitative features all the simulations we performed is shown in Fig. 1A. Starting with identical sequences folding into a random shape, the simulation was stopped when the population became dominated by the target, here a canonical tRNA shape. The black curve traces the average distance to the target (inversely related to fitness) in the population against time. Aside from a short initial phase, the entire history is dominated by steps, that is, flat periods of no apparent adaptive progress, interrupted by sudden approaches toward the target structure (*7*). However, the dominant shapes in the population not only change at these marked events but undergo several fitness-neutral transformations during the periods of no apparent progress. Although discontinuities in the fitness trace are evident, it is entirely unclear when and on the basis of what the series of successive phenotypes itself can be called continuous or discontinuous.

A set of entities is organized into a (topological) space by assigning to each entity a system of neighborhoods. In the present case, there are two kinds of entities: sequences and shapes, which are related by a thermodynamic folding procedure. The set of possible sequences (of fixed length) is naturally organized into a space because point mutations induce a canonical neighborhood. The neighborhood of a sequence consists of all its one-error mutants. The problem is how to organize the set of possible shapes into a space. The issue arises because, in contrast to sequences, there are

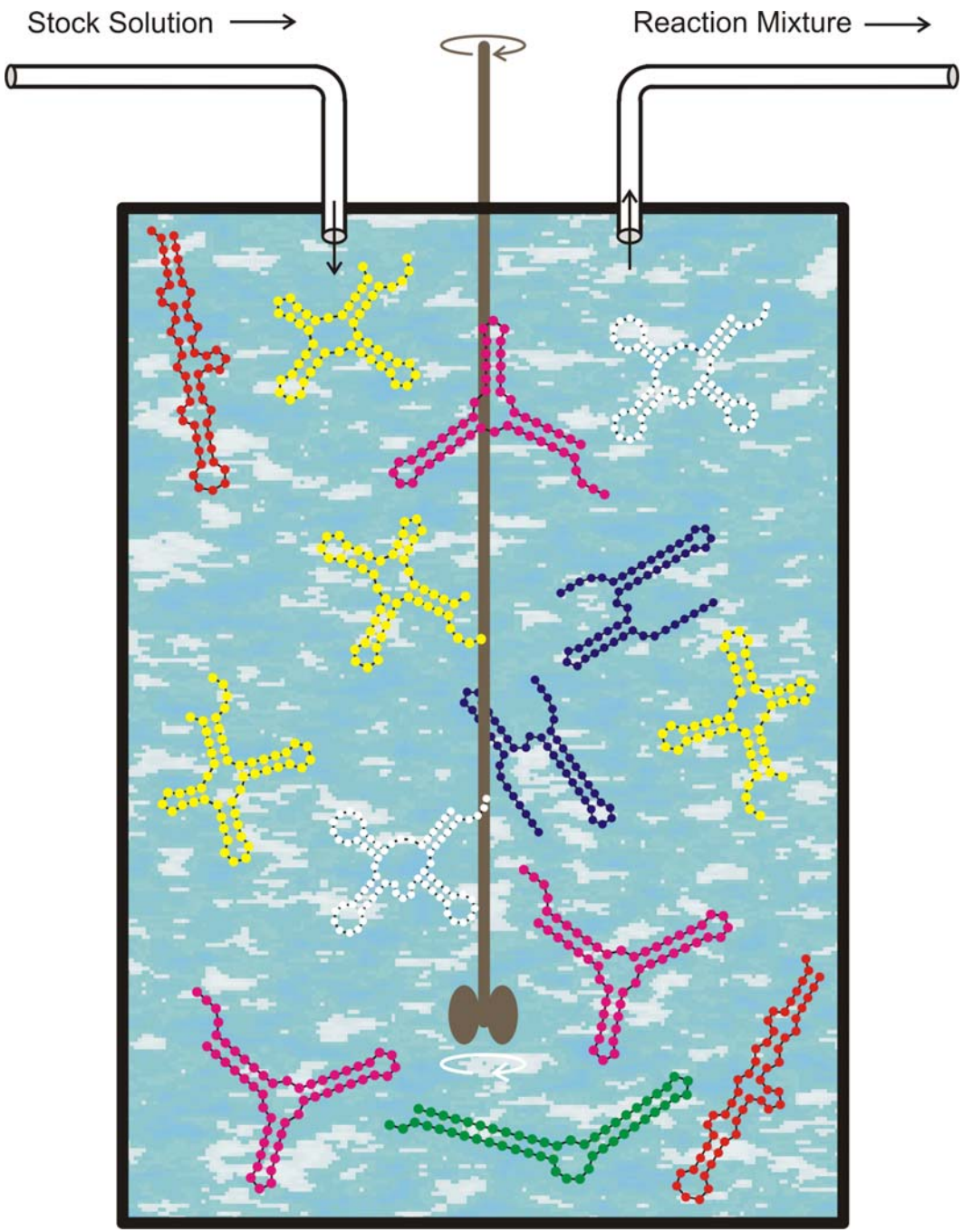
Institut für Theoretische Chemie, Universität Wien, Währingerstrasse 17, A-1090 Wien, Austria, Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA, and International Institute for Applied Systems Analysis (IIASA), A-2361 Laxenburg, Austria.




Structure of randomly chosen initial sequence




Phenylalanyl-tRNA as target structure




**Replication rate constant**

**(Fitness)**:

$$f_k = \gamma / [\alpha + \Delta d_S^{(k)}]$$

$$\Delta d_S^{(k)} = d_H(S_k, S_\tau)$$

**Selection pressure**:

The population size,

$N$ = # RNA moleucles,

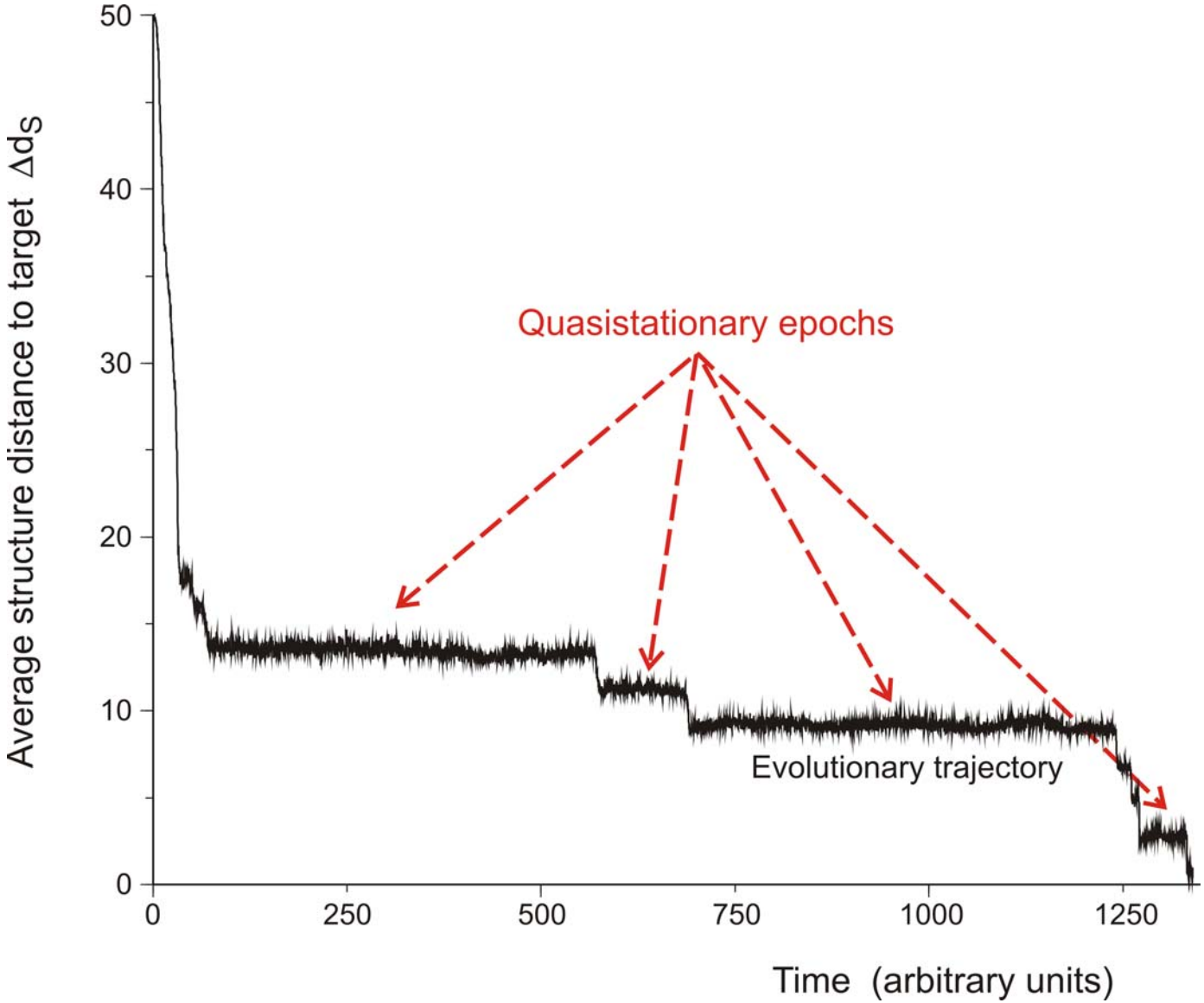
is determined by the flux:

$$N(t) \approx \overline{N} \pm \sqrt{\overline{N}}$$

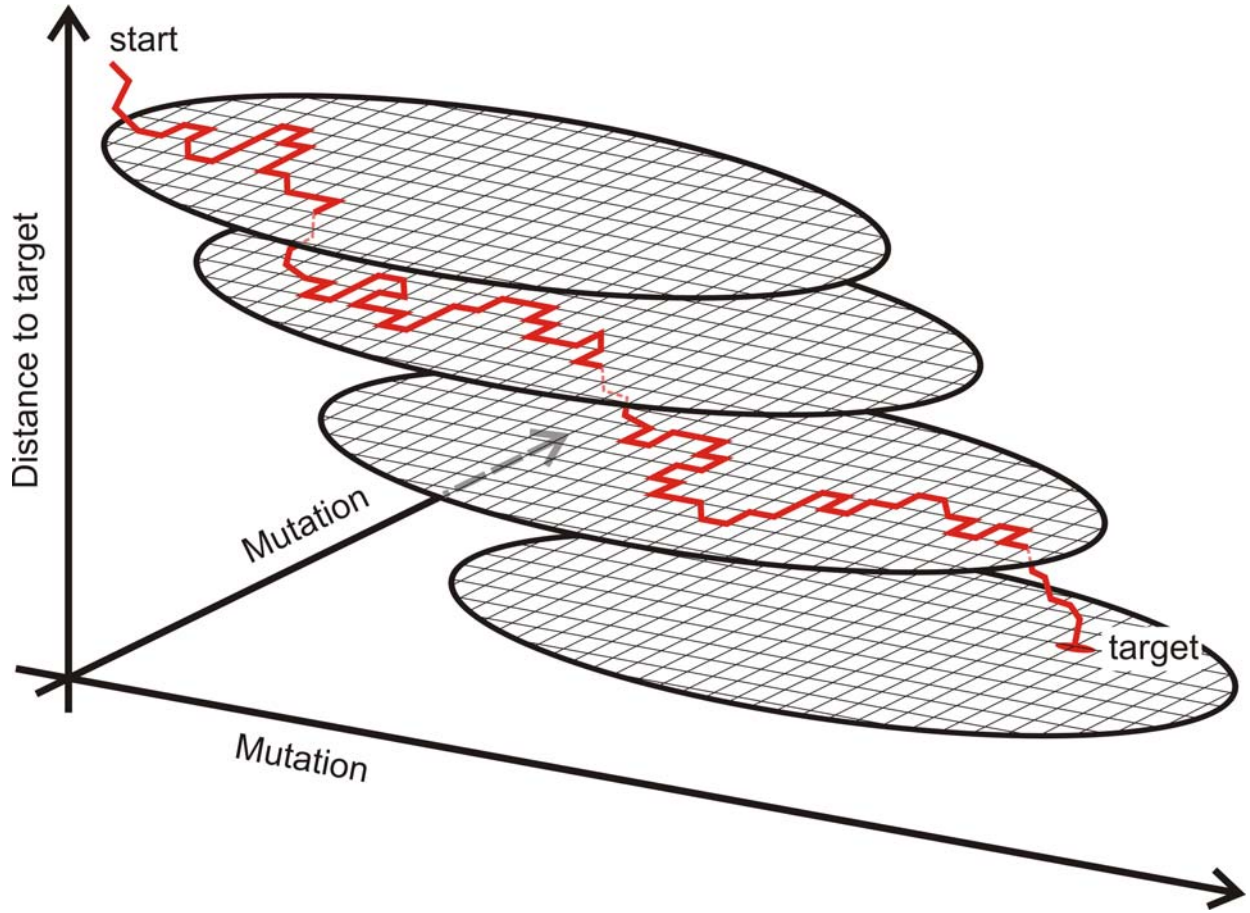
**Mutation rate**:

p = 0.001 / Nucleotide × Replication

The flow reactor as a device for studying the evolution of molecules ***in vitro*** and ***in silico***.

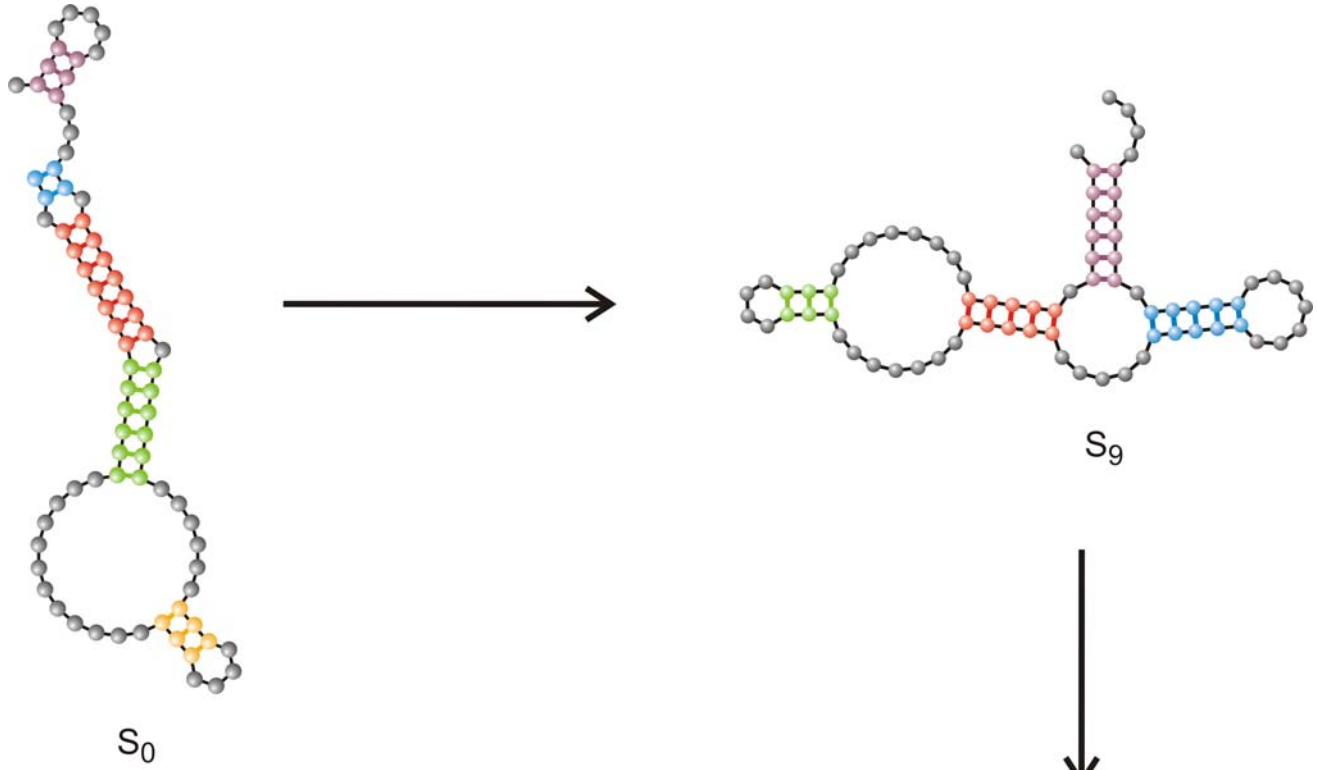



***In silico*** optimization in the flow reactor: Evolutionary Trajectory

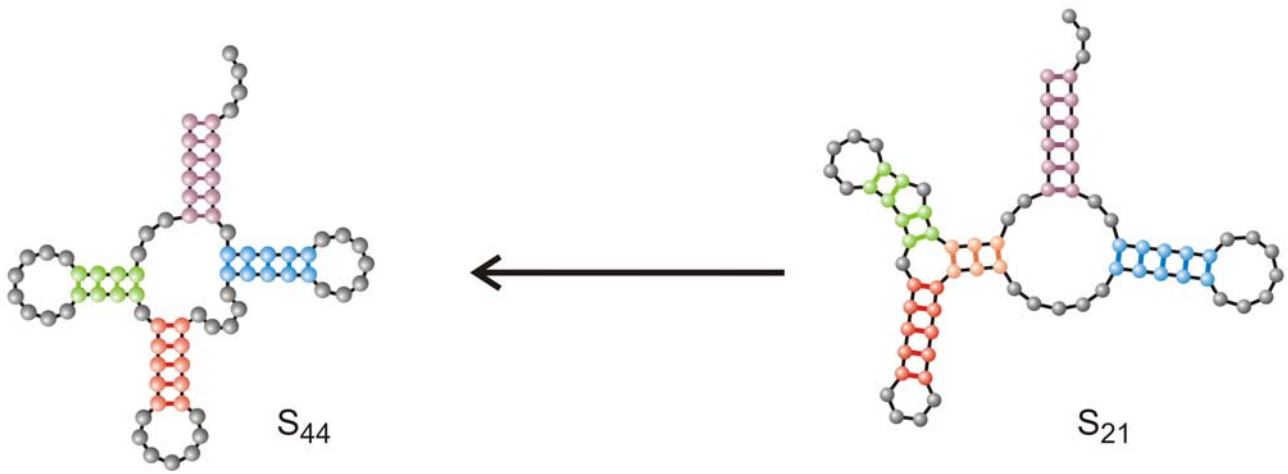



A sketch of optimization on neutral networks

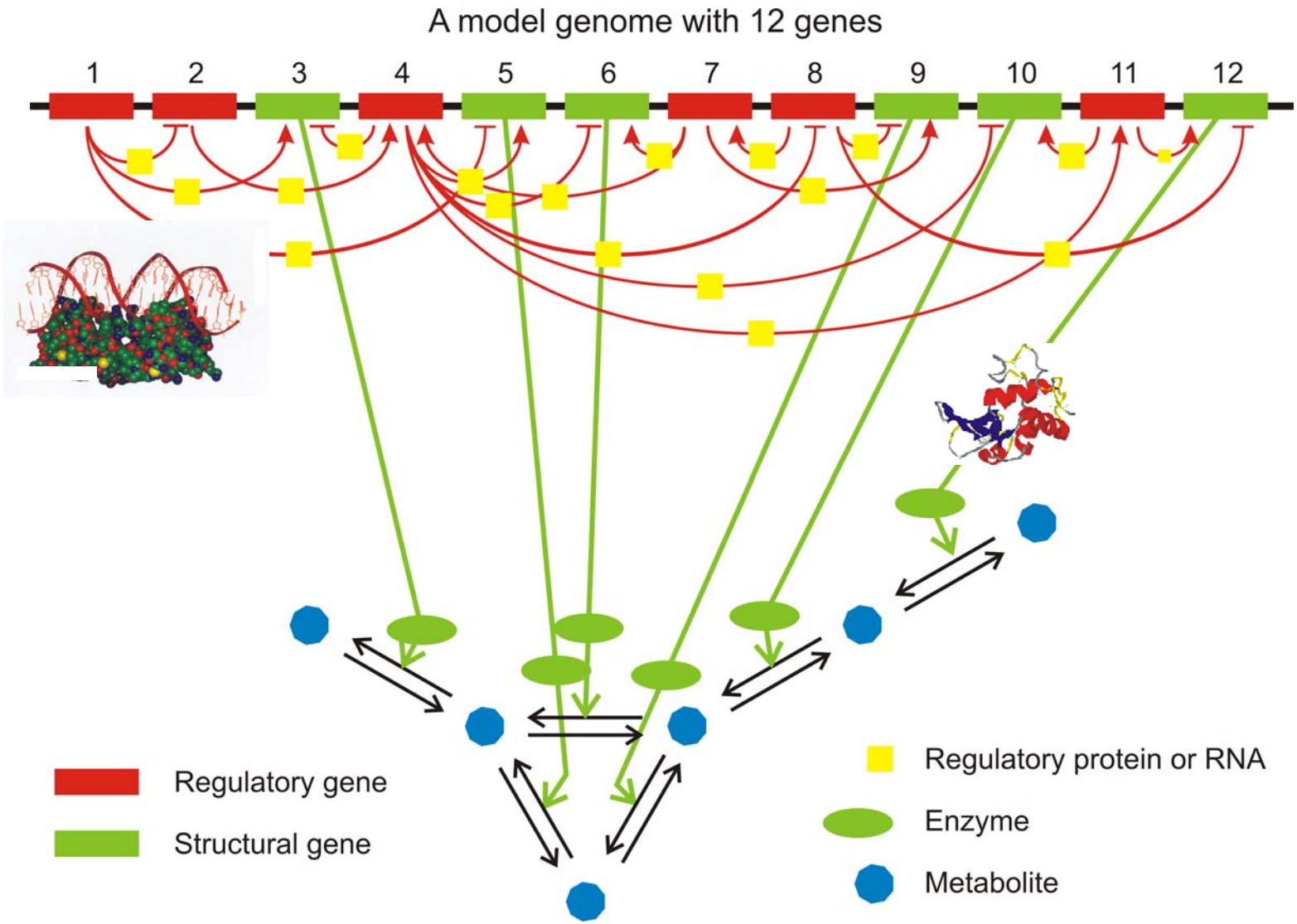
Randomly chosen
initial structure



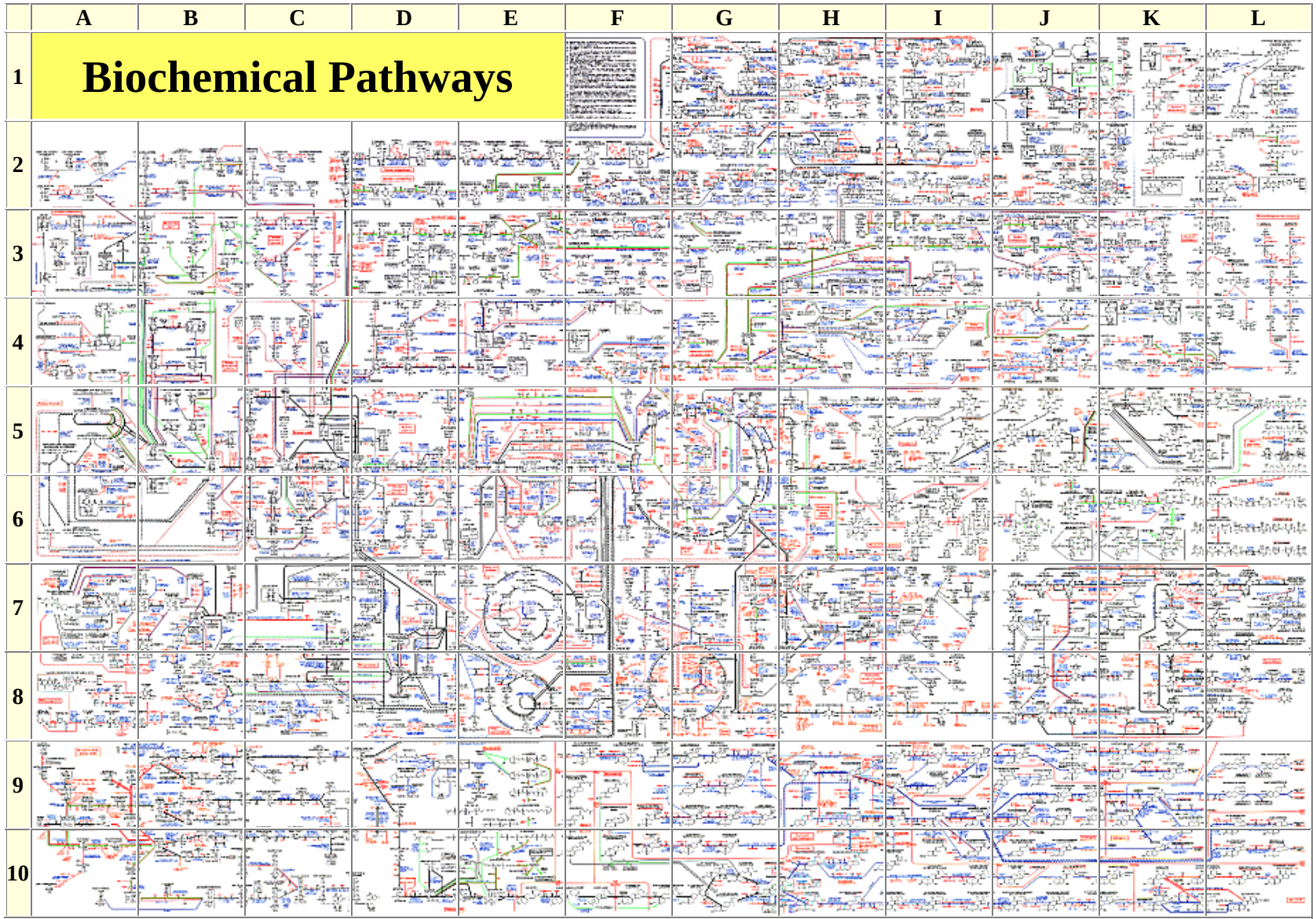

Phenylalanyl-tRNA
as target structure



1. Musterbildung in Physik und Chemie
2. Muster in der Biologie
3. Darwins Prinzip der natürlichen Auslese
4. Vermehrung und Evolution von Molekülen
5. Chemische Kinetik der molekularen Evolution
6. Evolutionsexperimente mit Molekülen
7. Ursachen und Konsequenzen der Neutralität
8. **Komplexität in der Biologie**

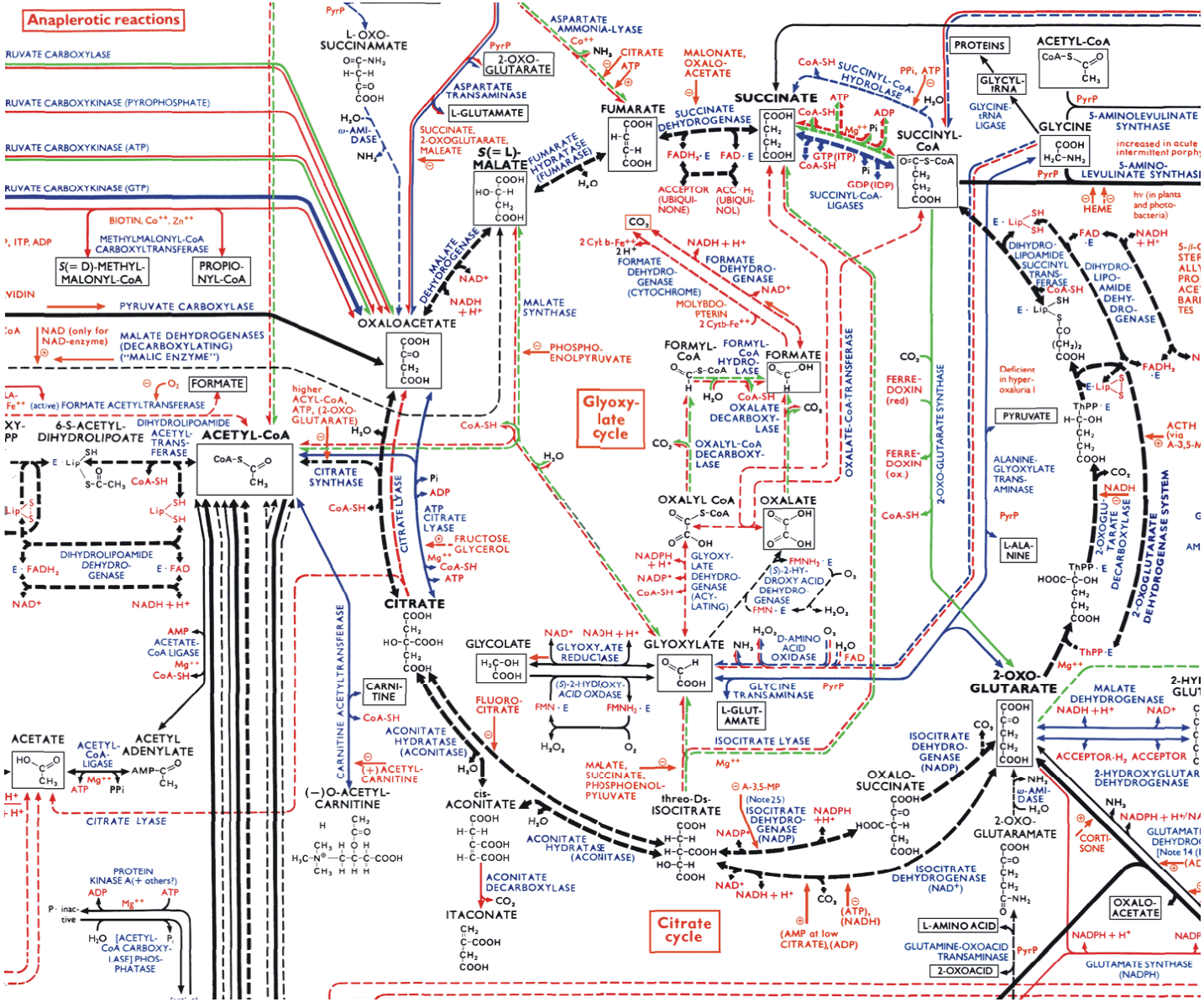



A sketch of a genetic and metabolic network




The reaction network of cellular metabolism published by Boehringer-Ingelheim.

The citric acid or Krebs cycle (enlarged from previous slide).

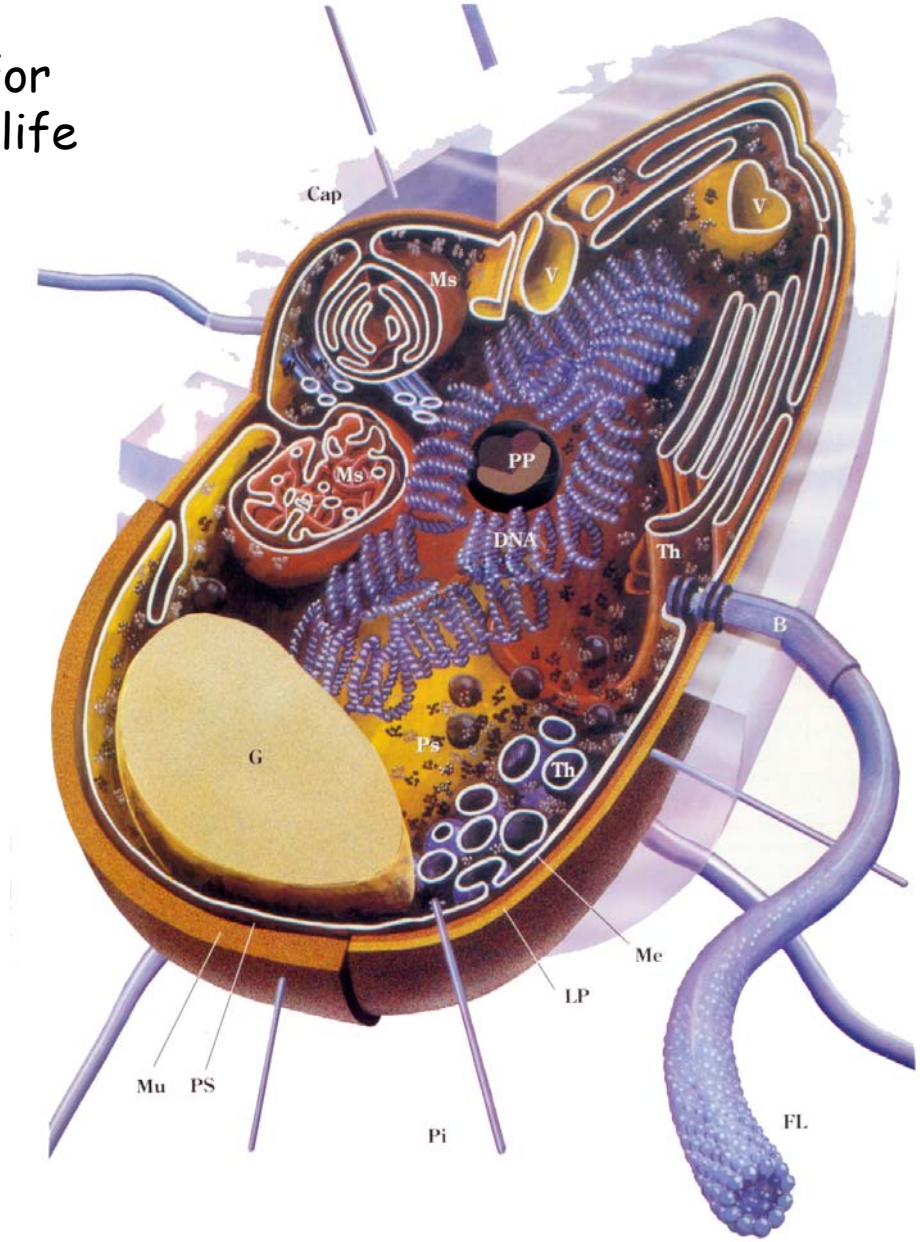


The bacterial cell as an example for the simplest form of autonomous life

Escherichia coli genome:

4 million nucleotides

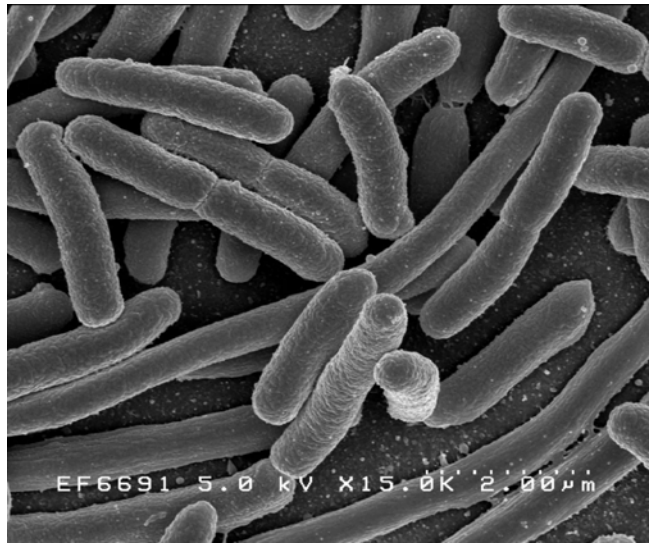
4460 genes




The structure of the bacterium *Escherichia coli*

| | | |
|---|---|---|
| **E. coli**: | Genome length | $4\times10^6$ nucleotides |
| | Number of cell types | 1 |
| | Number of genes | 4 460 |

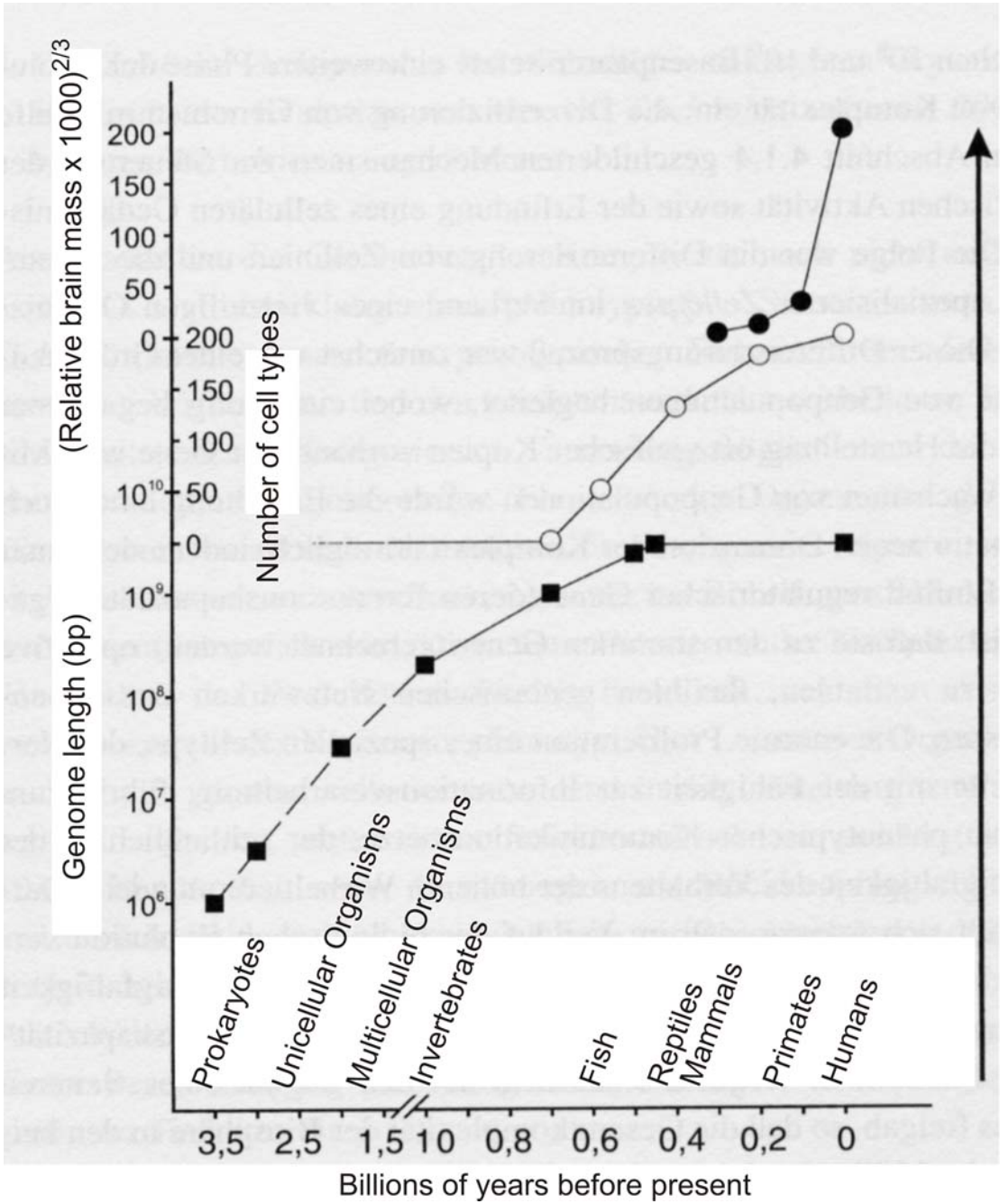
Four books, 300 pages each




| | | |
|---|---|---|
| **Man**: | Genome length | $3\times10^9$ nucleotides |
| | Number of cell types | 200 |
| | Number of genes | ≈ 30 000 |

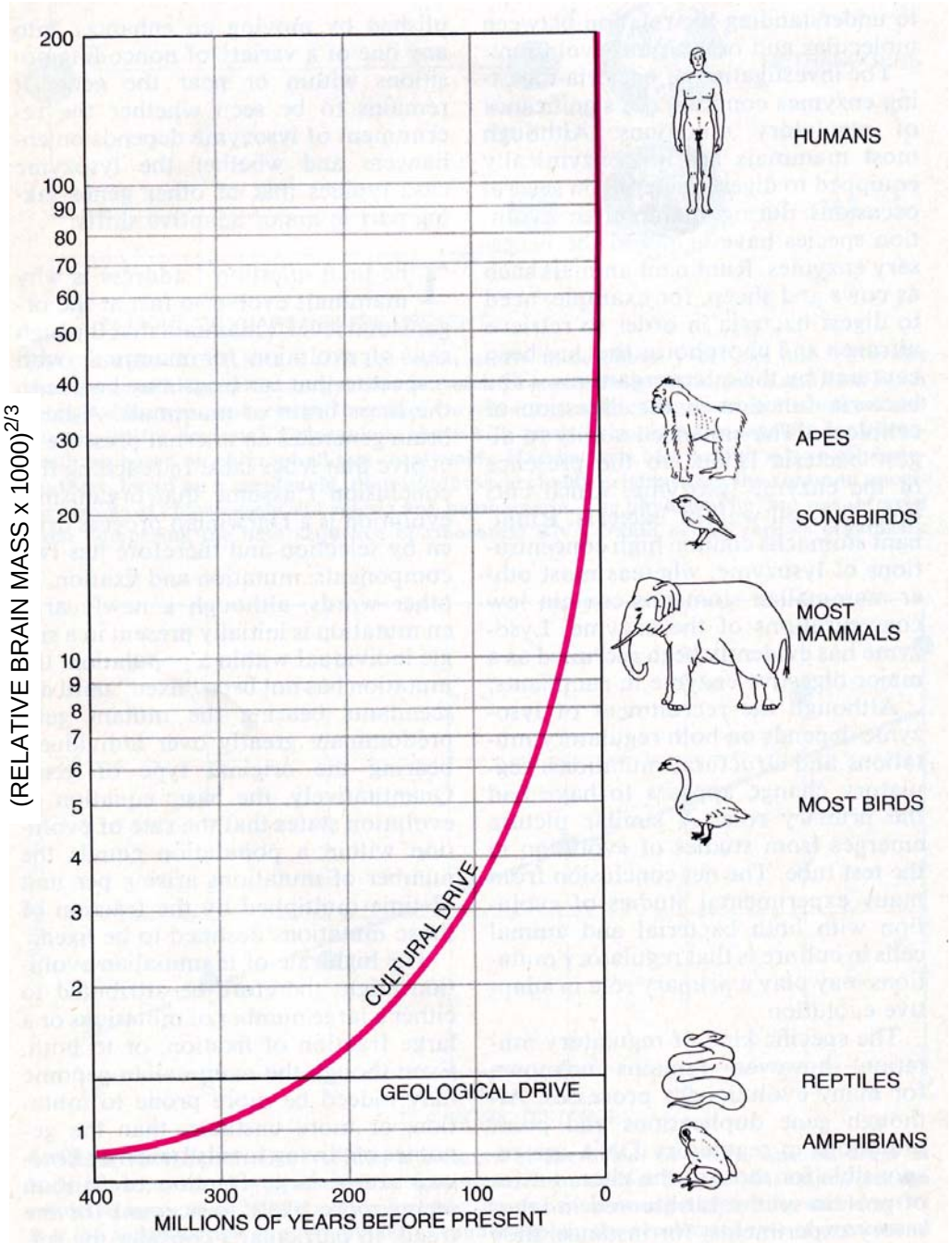
A library of 3000 volumes,
300 pages each

Complexity in biology




Wolfgang Wieser. 1998. ‚*Die Erfindung der Individualität*‘ oder ‚*Die zwei Gesichter der Evolution*‘. Spektrum Akademischer Verlag, Heidelberg 1998

BRITISH TIT




Alan C. Wilson.1985. The molecular basis of evolution.
*Scientific American* **253**(4):148-157.

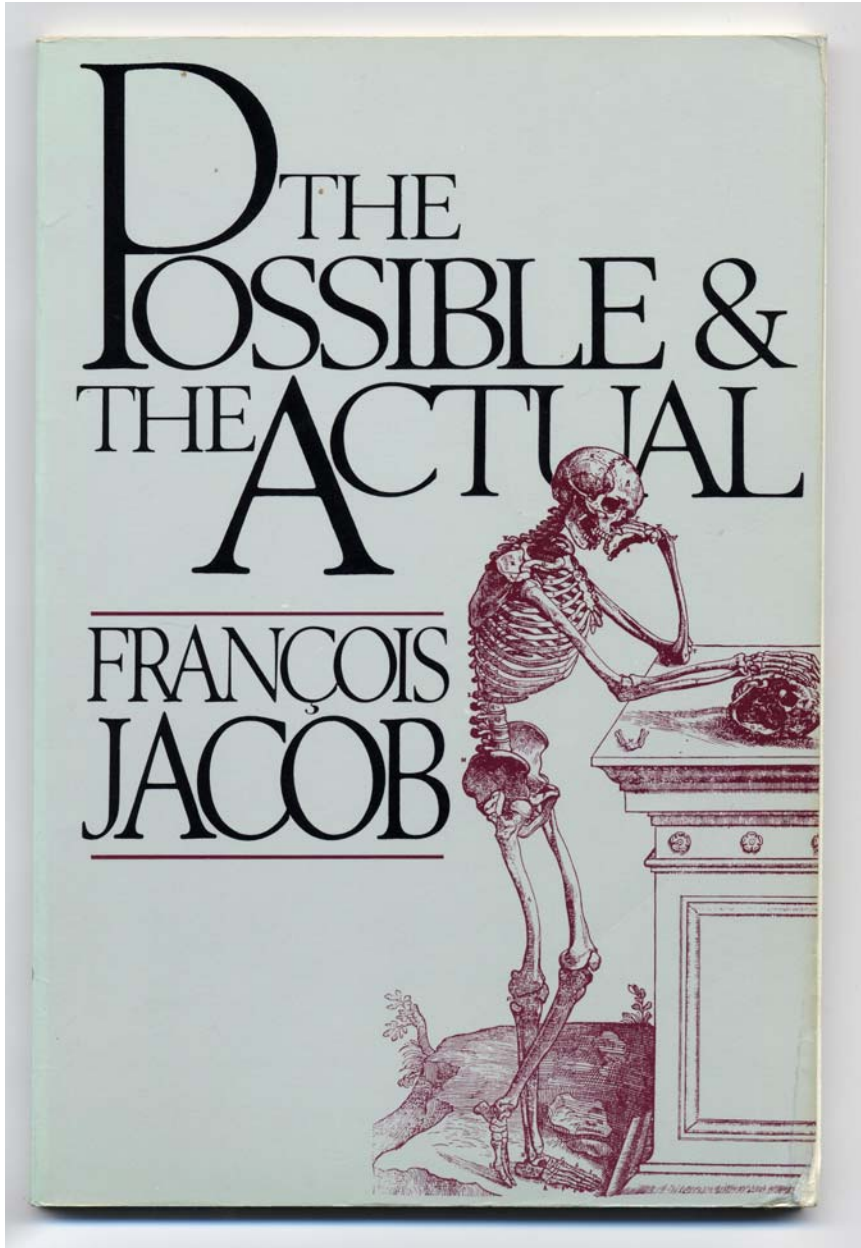
## *EVOLUTIONARY TINKERING*

*Blood . . . is the best possible thing to have coursing through one's veins.*

—*Woody Allen*, Getting Even

Evolution does not design with the eyes of an engineer, evolution works like a tinkerer.

Francois Jacob, Pantheon Books, New York 1982



The difficulty to define the notion of „gene".

Helen Pearson,
*Nature* **441**: 399-401, 2006



# WHAT IS A GENE?

The idea of genes as beads on a DNA string is fast fading. Protein-coding sequences have no clear beginning or end and RNA is a key part of the information package, reports **Helen Pearson**.

'Gene' is not a typical four-letter word. It is not offensive. It is never bleeped out of TV shows. And where the meaning of most four-letter words is all too clear, that of gene is not. The more expert scientists become in molecular genetics, the less easy it is to be sure about what, if anything, a gene actually is.

Rick Young, a geneticist at the Whitehead Institute in Cambridge, Massachusetts, says that when he first started teaching as a young professor two decades ago, it took him about two hours to teach fresh-faced undergraduates what a gene was and the nuts and bolts of how it worked. Today, he and his colleagues need three months of lectures to convey the concept of the gene, and that's not because the students are any less bright. "It takes a whole semester to teach this stuff to talented graduates," Young says. "It used to be we could give a one-off definition and now it's much more complicated."

In classical genetics, a gene was an abstract concept — a unit of inheritance that ferried a characteristic from parent to child. As biochemistry came into its own, those characteristics were associated with enzymes or proteins, one for each gene. And with the advent of molecular biology, genes became real, physical things — sequences of DNA which when converted into strands of so-called messenger RNA could be used as the basis for building their associated protein piece by piece. The great coiled DNA molecules of the chromosomes were seen as long strings on which gene sequences sat like discrete beads.

This picture is still the working model for many scientists. But those at the forefront of genetic research see it as increasingly old-fashioned — a crude approximation that, at best, hides fascinating new complexities and, at worst, blinds its users to useful new paths of enquiry.

Information, it seems, is parceled out along chromosomes in a much more complex way than was originally supposed. RNA molecules are not just passive conduits through which the gene's message flows into the world but active regulators of cellular processes. In some cases, RNA may even pass information across generations — normally the sole preserve of DNA.

An eye-opening study last year raised the possibility that plants sometimes rewrite their DNA on the basis of RNA messages inherited from generations past[1]. A study on page 469 of this issue suggests that a comparable phenomenon might occur in mice, and by implication in other mammals[2]. If this type of phenomenon is indeed widespread, it "would have huge implications," says evolutionary geneticist Laurence Hurst at the University of Bath, UK.

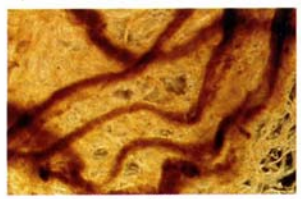
"All of that information seriously challenges our conventional definition of a gene," says molecular biologist Bing Ren at the University of California, San Diego. And the information challenge is about to get even tougher. Later this year, a glut of data will be released from the international Encyclopedia of DNA Elements (ENCODE) project. The pilot phase of ENCODE involves scrutinizing roughly 1% of the human genome in unprecedented detail; the aim is to find all the sequences that serve a useful purpose and explain what that purpose is. "When we started the ENCODE project I had a different view of what a gene was," says contributing researcher Roderic Guigo at the Center for Genomic Regulation in Barcelona. "The degree of complexity we've seen was not anticipated."

**"We've come to the realization that the genome is full of overlapping transcripts."**
— Phillip Kapranov

## Under fire

The first of the complexities to challenge molecular biology's paradigm of a single DNA sequence encoding a single protein was alternative splicing, discovered in viruses in 1977 (see 'Hard to track', overleaf). Most of the DNA sequences describing proteins in humans have a modular arrangement in which exons, which carry the instructions for making proteins, are interspersed with non-coding introns. In alternative splicing, the cell snips out introns and sews together the exons in various different orders, creating messages that can code for different proteins. Over the years geneticists have also documented overlapping genes, genes within genes and countless other weird arrangements (see 'Muddling over genes', overleaf).

Alternative splicing, however, did not in itself require a drastic reappraisal of the notion of a gene; it just showed that some DNA sequences could describe more than one protein. Today's assault on the gene concept is more far reaching, fuelled largely by studies that show the previously unimagined scope of RNA.



Spools of DNA (above) still harbour surprises, with one protein-coding gene often overlapping the next.

P. PLAILLY/SPL

The one gene, one protein idea is coming under particular assault from researchers who are comprehensively extracting and analysing the RNA messages, or transcripts, manufactured by genomes, including the human and mouse genome. Researchers led by Thomas Gingeras at the company Affymetrix in Santa Clara, California, for example, recently studied all the transcripts from ten chromosomes across eight human cell lines and worked out precisely where on the chromosomes each of the transcripts came from[3].

The picture these studies paint is one of mind-boggling complexity. Instead of discrete genes dutifully mass-producing identical RNA transcripts, a teeming mass of transcription converts many segments of the genome into multiple RNA ribbons of differing lengths. These ribbons can be generated from both strands of DNA, rather than from just one as was conventionally thought. Some of these transcripts come from regions of DNA previously identified as holding protein-coding genes. But many do not. "It's somewhat revolutionary," says Gingeras's colleague Phillip Kapranov. "We've come to the realization that the genome is full of overlapping transcripts."

Other studies, one by Guigo's team[4], and one by geneticist Rotem Sorek[5], now at Tel Aviv University, Israel, and his colleagues, have hinted at the reasons behind the mass of transcription. The two teams investigated occasional reports that transcription can start at a DNA sequence associated with one protein and run straight through into the gene for a completely different protein, producing a fused transcript. By delving into databases of human RNA transcripts, Guigo's team estimate that 4–5% of the DNA in regions conventionally recognized as genes is transcribed in this way. Producing fused transcripts could be one way for a cell to generate a greater variety of proteins from a limited number of exons, the researchers say.

Many scientists are now starting to think that the descriptions of proteins encoded in DNA know no borders — that each sequence reaches into the next and beyond. This idea will be one of the central points to emerge from the ENCODE project when its results are published later this year.

Kapranov and others say that they have documented many examples of transcripts in which protein-coding exons from one part of the genome combine with exons from another

ENCODE stands for
ENCyclopedia Of DNA Elements.

**ENCODE** Project Consortium.
Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project.
*Nature* **447**:799-816, 2007

Web-Page for further information:

http://www.tbi.univie.ac.at/~pks